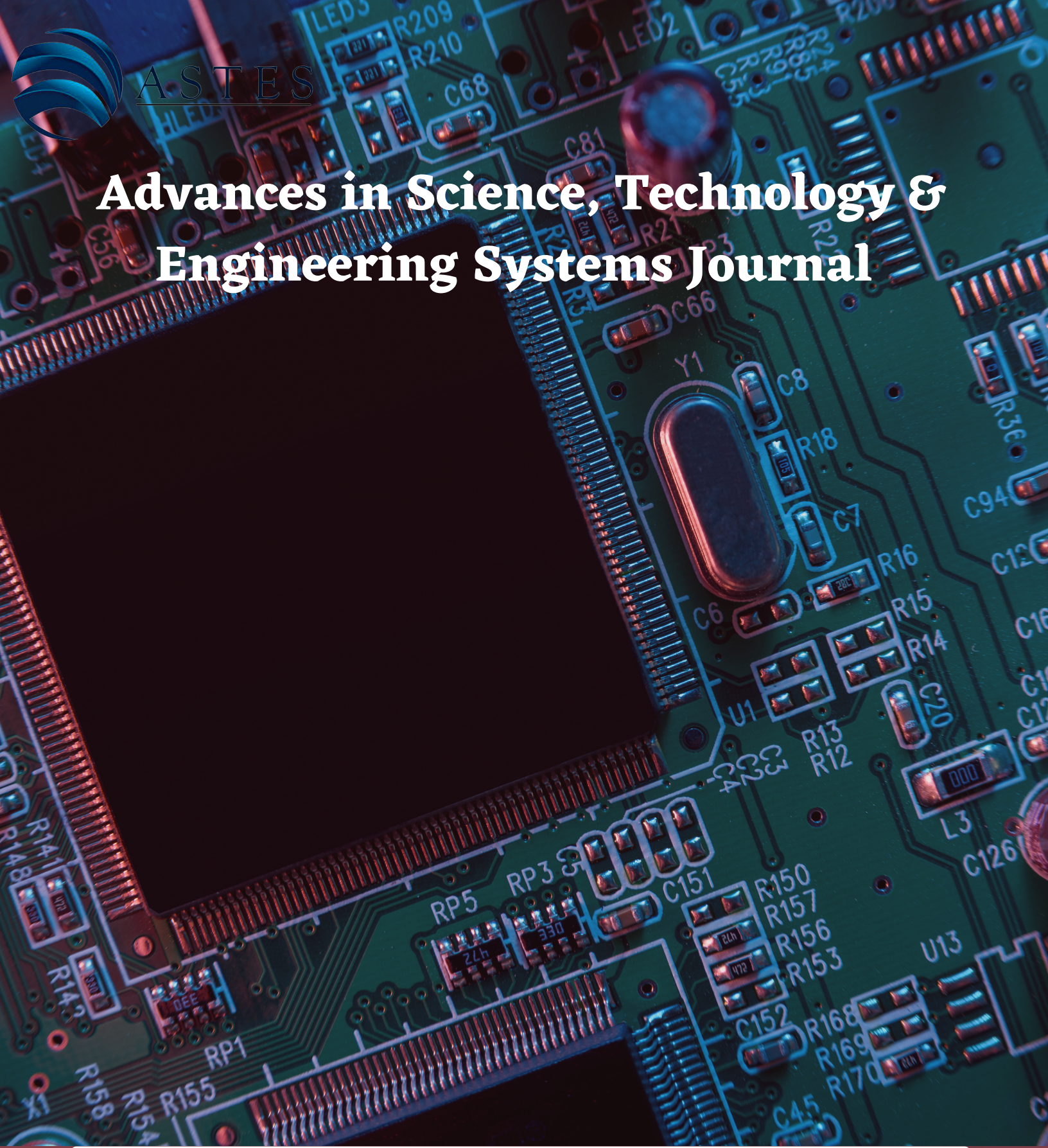


Advances in Science, Technology & Engineering Systems Journal

VOLUME 5-ISSUE 1|JAN-FEB 2020

www.astesj.com
ISSN: 2415-6698

Editorial

Advances in Science, Technology and Engineering Systems Journal (ASTESJ) is an online-only journal dedicated to publishing significant advances covering all aspects of technology relevant to the physical science and engineering communities. The journal regularly publishes articles covering specific topics of interest.

Current Issue features key papers related to multidisciplinary domains involving complex system stemming from numerous disciplines; this is exactly how this journal differs from other interdisciplinary and multidisciplinary engineering journals. This issue contains 57 accepted papers in Computer Science and Electronics domains.

Editor-in-chief
Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 5 Issue 1 **January-February 2020**

CONTENTS

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 01-07 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Single-leg Standing Ability and Lower Limb Movement Analysis of Collegiate Footballers and Sedentary Students

Jiun Sien Lau, Vina Phei Sean Tan, Mohd Akmal Vahidi, Siti Hanisah Linggi Mohd Azam, Mohd Hafezi Mat Zain, Azilawati Azizan, Mazra Othman, Rosniwati Ghafar*

Exercise and Sports Science, School of Health Sciences, Universiti Sains Malaysia, Health Campus, 16150, Malaysia

ARTICLE INFO

Article history:
Received: 30 September, 2019
Accepted: 10 December, 2019
Online: 15 January, 2020

Keywords:
Balancing
Lower limbs
Footballer
Sedentary

ABSTRACT

Stability or balance is an integral component to perform daily activities without incurring injury or to be dependent on others. Sportsmen tend to have better balance than non-sports people but less is known about the single-leg balance ability. Furthermore, few studies analyzed the dynamic phase of single-leg stance that may contribute to better overall balance. Sports like football tend to have instances where the player's non-dominant leg keeps them in an upright position while the dominant leg kicks, passes and stops the ball. We aim to study the single-leg balance between collegiate footballers and sedentary students in eyes-closed (EC) and eyes-opened (EO) conditions and their contributing components to keep a body in an up-right position. Twenty collegiate footballers and 20 sedentary students conducted the unipedal stance test (UPST) on each leg with EO and EC conditions while standing on a force platform. We captured center of pressure (CoP) distance travelled, stance duration and using a 3D motion capture system, we assessed lower limb movement at six different anatomical sites. Results showed that footballers had better overall balance compared to sedentary students only in the non-dominant leg, EC condition with 12 footballers versus four sedentary students completing the full 45s stance ($p=0.01$) The other three UPST conditions did not show significant differences between groups. The CoP distance in the initial dynamic state and total UPST were both significantly shorter in footballers than sedentary students ($p<0.05$) during the non-dominant leg, EC stance. Our multivariable linear regression model significantly predicted time for UPST on non-dominant leg with EC up to 76.8% ($p<0.001$) with the first 5-s of greater trochanter movement significantly contributing to total time taken for UPST in footballers. Overall, playing football may enhance balance control intrinsically especially for the non-dominant side while being less reliant on visual input to maintain balance.

1. Introduction

Balance is a crucial motor skill stemming from muscle synergies that lessens centre of pressure (CoP) displacement to effectively keep our body upright in a standing position, during locomotion and body orientation [1]. These muscle synergies are coordinated by the central nervous system with various inputs from our visual, vestibular and somatosensory systems [2]. In sports like gymnastics, balance is key to excel in the sport, however, other sports like football and basketball, static and dynamic balance are just as important although less apparent in their movements [3].

In football (considered as soccer in North America), most players shoot, pass and stop the ball with their dominant leg while the non-dominant leg supports their body weight [4]. Footballers often play and train on various turf conditions and wear cleated or non-cleated shoes. These conditions challenge trained footballers to develop better balance while mastering their football skills. Hence, top performing footballers have better posture stability and use less visual information to maintain their balance [5].

It is important to assess balance ability as it may relate to the level of sports performance and injury prevention. Single-limb and two-limb stance tests with eyes opened and with eyes closed are commonly used to test standing balance [6]. The unipedal stance

*Corresponding Author: Rosniwati Ghafar, rosnikk@usm.my

test (UPST) is a common balance and standing ability assessment that can be easily performed and does not require additional equipment [7]. However, UPST lacks normative values for different age groups and balance ability changes with age [8]. Yet, UPST could detect changes of postural balance after 12 weeks of recreational soccer training and running in untrained men [9]. Researchers concluded that soccer training and high-intensity interval running enhanced the balance ability in their study participants [9]. As the UPST mainly measures the time taken to perform the test, we can also assess the center of pressure (CoP) during UPST using advanced technology to further understand the mechanisms of balance.

There are various types of force or pressure assessment systems to quantify CoP. Force platforms detect the ground reaction force and CoP. CoP can also be assessed from the force distribution across a sensor grid by other pressure assessment systems [10].

In sports biomechanics, motion data are used to study and observe human performance. The three-dimensional (3-D) motion capture system consists of infrared cameras with reflective markers attached to anatomical bony landmarks [11]. This system captures and measures motion that was hard to assess previously. 3-D motion capture plays an important role in sports to analyze physical limitations and for movement optimization [11]. Through the analyses, details about injury mechanism and movement will lead to new preventive methods and be used to improve a player's technique and performance [12].

There are two distinct phases in UPST based on Jonsson and colleagues' research [13]. The first phase is the first 5 s of UPST and also known as the dynamic phase where a rapid reduction in force variability can be seen. The better control a person has in the dynamic phase, the better their balance will be to complete the 45-s UPST. The second phase is known as the static phase where the individual maintains a certain level of force variability that enables balance to be achieved [13]. Jonsson and colleagues concluded that the dynamic phase was key in assessing UPST [13]. Yet, we know little about lower limb influence on single-leg stance in young adults [14].

To our knowledge, studies that compared the balance ability between collegiate footballers and sedentary students are limited despite the possibility of recreational football play could enhance balance ability in previously untrained men [9]. Furthermore, balance is an important functional fitness component and should be cultivated to avoid balance-related injuries [15]. Therefore, our study aims to investigate if indeed simple football play could be related to better balance in young adults. We will compare the standing ability between collegiate footballers and sedentary students. We will also conduct further analysis on the dynamic phase of UPST to better understand lower limb control during the single-leg balance test.

2. Methodology

2.1. Participants

We recruited male collegiate students aged between 18 to 25 years old and grouped them as footballers (n= 20) or sedentary students (n=20) based on our study's participant criteria. Footballers need to have a minimum of 2 years of football competition experience at collegiate level and/or exercise at least 3 times or more a week. Footballers need not be part of any specific training other than their recreational football games during non-competition time. Sedentary students were those who exercised less than twice a week and do not participate in any sport competition. We excluded individuals with health conditions that may affect balance such as diabetes mellitus, musculoskeletal disorders, vestibular impairments and those with any lower limb injuries in the past 6 months. All participants provided informed consent and the Human Research Ethical Committee of Universiti Sains Malaysia provided ethical approval for this study.

2.2. Study Procedure

We measured participants' height (cm) and weight (kg) using a stadiometer and digital weight scale, respectively. Measures were done in duplicates and a third measure was obtained if difference between two measures were 0.4 cm in height or 0.2 kg in weight. We used the averaged value of measures obtained and calculated body mass index (BMI, kg/m^2) of all participants.

Participants conducted the UPST for both dominant and non-dominant legs with i) eyes-opened (EO) and ii) eyes-closed (EC) conditions. Participants were instructed to kick a ball and the leg used was considered the dominant leg. Test was terminated early if participants i) used arms to balance their body, ii) touched the floor with the raised foot, iii) moved the standing foot to maintain their body posture, such as rotating the foot on the force platform or iv) opened their eyes in the EC condition. [8]. Each participant conducted three UPST trials and had 30 s of rest between trials. We recorded the total time taken to maintain the single-leg stance and stopped the UPST after 45 s.

Participants stood on a force platform (Bertec, USA) and perform UPST by raising one leg as shown in Figure 1. The sampling rate of the platform was set at 120 Hz. Participants stood barefooted on the level platform with the foot positioning marked on the force platform. This ensures uniformity and standardization of feet placement on the force platform which minimise variations between trials [16].

In the attempts to maintain posture in the UPST, body muscles are constantly adjusting and this creates 'sway' or body movement that can be captured by the force platform [17]. The force platform captures CoP in the x- and y-axis in a constant flow and consequently, we assessed the distance travelled by CoP. Lesser distance and smaller CoP range usually indicates better balance skills [18].

A 3D motion capture system (Qualysis AB, Gothenburg Sweden) recorded lower limb movement and time during UPST. Six motion analysis cameras were placed surrounding the force platform as shown in Figure 2. In order to record movement of the limb during UPST, we attached 12 reflective markers to specific anatomical bony landmarks on the lower extremities. The reflective markers were attached to the greater trochanter, lateral side of the knee condyle, ankle and heel, and at the first and fifth metatarsal of both legs. The motion camera system captured the movement made during different UPST conditions and CoP from the force platform. Total distance was calculated from the movement of COP during performing UPST. The average time of the three trials and coordinates of CoP were extracted [8].

The first five seconds of UPST were further analysed to understand more about the dynamic phase of UPST. We used the results of the best trial out of the three attempts in this analysis. The best trial was selected based on the duration of the UPST and the shortest distance of CoP travelled during the same UPST trial.

2.3. Statistical Analysis

Statistical analyses were conducted using SPSS version 24.0 with significance set at p-value <0.05. Time taken to complete UPST had a skewed distribution and we used Mann-Whitney to compare the differences between groups. We used independent t-test to compare the differences between footballers and sedentary students for the normally distributed data from different UPST conditions.

To understand the relation of lower limb movement on UPST, we conducted Pearson's correlation tests to identify the related anatomical sites during the dynamic phase that contributes to the time taken to complete UPST. Subsequently, we included relevant variables in a multivariable linear regression analysis.

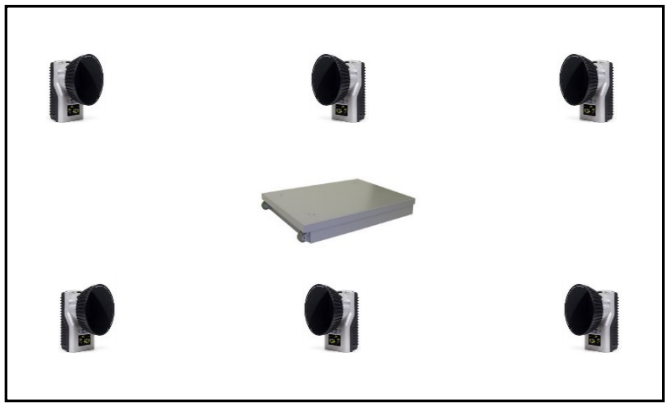


Figure 2: The field set-up of infrared cameras relative to the force platform

3. Results

Footballers were significantly younger than sedentary students by about 14 months but were similar in build (height, weight and BMI) as displayed in Table 1.

Table 1. Study participant characteristics in mean±standard deviation.

	Football (n=20)	Sedentary (n=20)	*p*-value
Age (yr)†	21.4 ± 1.1	22.6± 1.2	0.002
Height (m)	1.69±0.07	1.66±0.05	0.089
Weight (kg)	66.3±14.0	60.7±14.0	0.056
BMI (kg/m^2)	23.2±4.2	22.3±5.8	0.099

†*Differences between groups analysed using independent t-test while other variables were analysed using Mann-Whitney test.*
BMI, body mass index.

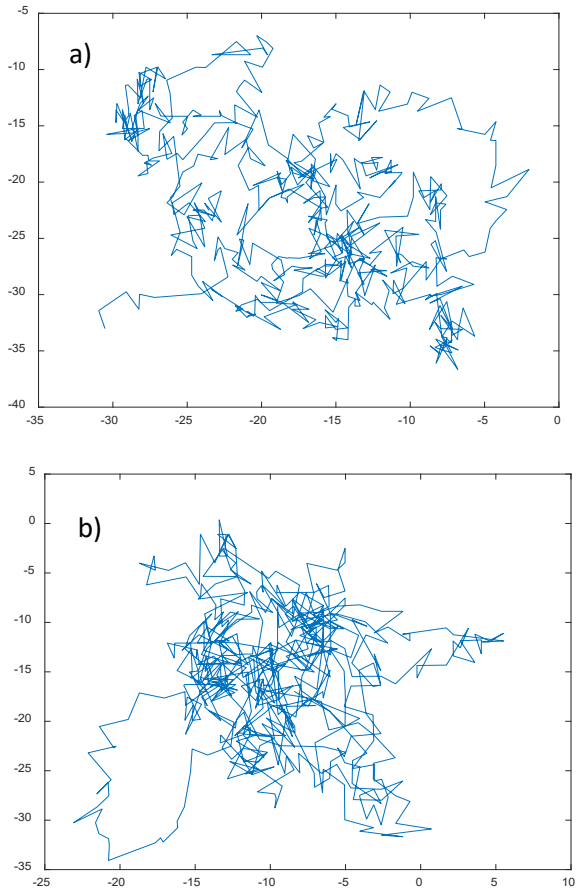


Figure 3: Examples of the 'best' or smallest range of CoP travel pathway for a) footballer and b) sedentary student during the non-dominant, eyes-closed condition.

3.1. Time Taken to Perform UPST

All footballers had no issues completing the UPST at the maximal 45 s when eyes were opened while standing on their dominant foot and only one person did not make the 45 s mark when standing on their non-dominant leg with EO (Table 2). Sedentary students did not perform any different than footballers on their dominant leg with EO or EC. However, there were three times more footballers (n=12) that could complete UPST till 45 s compared to sedentary students (n=4) when performing on their non-dominant leg with EC (p=0.01).

As data was not normally distributed in the time taken to perform UPST in four different conditions, we did not detect any significant differences (p>0.05) in time between groups as tested using the Mann-Whitney test.

3.2. Distance Travelled of CoP

Footballers were no different in their total CoP distance travelled during UPST than sedentary students using their dominant leg in both EO and EC conditions. From Table 2, total CoP distance on non-dominant leg, EO UPST were no different between groups too ($p>0.05$). In non-dominant leg, EC UPST, footballers significantly swayed less compared to sedentary students ($p=0.01$).

3.3. Distance Travelled of CoP in Dynamic Phase

Examples of CoP patterns of a footballer versus a sedentary student are presented in Figure 3 and 4 during the dynamic phase or the first 5-s portion of UPST. We selected the best (smallest) and worst (largest) range of CoP for both groups respectively in the non-dominant leg, EC session. Footballer CoPs showed less complexity of pattern within the range compared to a sedentary students' CoP travel path.

The 5-s CoP distance travelled were thus significantly different ($p=0.021$) between footballers and sedentary students when analysed using Mann-Whitney test (Table 2). The rest of the conditions were no different in 5-s CoP distance reported between footballers and sedentary students.

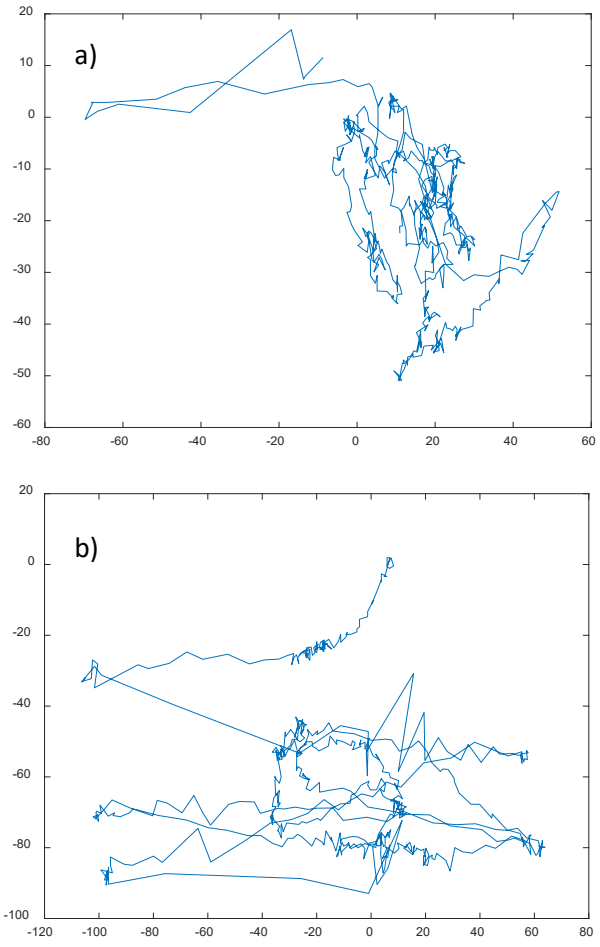


Figure 4: Example of a 'worst' or largest range of CoP travel pathway for a) footballer and b) sedentary student during the non-dominant, eyes-closed condition.

3.4. Non-dominant Leg, EC UPST Time and Lower Limb Control During Dynamic Phase

In Table 3, we used the best trial to further study the contribution of lower limb movement to overall time taken to complete UPST. Footballers could maintain their UPST significantly longer than sedentary students using the non-dominant leg in the EC condition.

We analysed the first 5-s of lower limb function on single-leg balance during the dynamic phase toward overall time taken to complete UPST. Interestingly, the lower limb movement in all six anatomical sites were not significantly different between footballers and sedentary students (Table 3). Yet, we observed significant negative correlation of all lower limb movement sites to time taken to complete UPST (Table 4). However, we had to remove four data sets due to noise in lower limb movement analysis.

Subsequently, we included the lower limb movements in the first 5-s of UPST to predict time to complete UPST. We analysed footballers and sedentary students separately. We found that different sites of lower limb movement were not predictive of overall time to complete UPST in sedentary students ($p>0.05$). However, our regression model in footballers (Table 5) showed that the overall model significantly predicted up to 76.1% ($p<0.001$) of total time to complete UPST. In that regression model, although we have ankle, knee and greater trochanter as predictor variable, the movement from the greater trochanter was the significant variable to predict time to complete UPST. The first and fifth metatarsal and heel sites were removed due to high correlation within the regression model (variance inflation factor >5).

4. Discussion

The objective of the study is to compare the standing balance ability between male collegiate football players and sedentary students and to understand the influence in the dynamic phase toward overall single-leg standing. We found that footballers on non-dominant leg with eyes-closed (EC) could significantly balance better than sedentary students as they completed more trials, had a longer balance time and CoP distance was shorter in the dynamic and static phase. Furthermore, we found that the footballers' greater trochanter control in the early dynamic phase significantly predicts overall balance performance in the non-dominant leg, EC condition. Whereas lower limb movements of sedentary students may not be helpful in maintaining balance during an EC, single-leg stand.

Differences in balance ability between sports/trained with untrained individuals are known. Specifically, in football, two other studies have showed that the highly skilled footballers relied less on visual information to maintain a stable posture [5,19]. Naturally, footballers were also known to be proficient in using their non-dominant leg as well as their dominant side [20]. Studies of other sports that included gymnastics and swimming also showed better postural control and efficient use of somatosensory and otolithic input to maintain body posture [21, 22]. Results of sports contributing to better balance is even apparent in recreational players whereby simple football training resulted in adapted visual, somatosensory and vestibular senses that improved

Table 2. Outcomes of number of individuals that completed UPST till 45 s and the average of three trials for i) time taken to complete unipedal stance test (UPST), ii) total CoP distance in complete UPST and iii) CoP distance in the first 5-sec of UPST in footballers and sedentary students, reported as median (inter-quartile range).

	Footballer (n=20)				Sedentary students (n=20)			
	Dominant leg		Non-dominant leg		Dominant leg		Non-dominant leg	
	EO	EC	EO	EC	EO	EC	EO	EC
Completed max. 45-s UPST (n)	20	11	19	12	19	9	18	4*
Duration of UPST (s)	45.0 (0.0)	23.0 (32.6)	45.0 (0.0)	31.2 (27.1)	45.0 (0.0)	20.4 (21.6)	45.0 (13.7)	24.4 (21.8)
Total CoP distance (mm)†	144.8 (2.6)	140.1 (3.9)	144.3 (3.0)	140.5 (3.9)	145.8 (0.6)	140.8 (5.0)	145.8 (0.6)	144.1 (4.7)*
CoP distance in first 5-s (mm)	133.8 (0.1)	134.0 (0.7)	133.8 (0.1)	134.3 (0.5)	133.8 (0.1)	134.1 (0.3)	133.8 (0.1)	134.0 (0.3)*

EO, eyes-opened; EC, eyes-closed.
**Significant differences between footballers and sedentary students in similar leg and eye conditions with p-value <0.05 as analysed by Mann-Whitney test or independent t-test.*
† Differences between groups were analysed using independent t-test and results reported as mean and standard deviation.

Table 3. Comparison of time, centre of pressure (CoP) range of x- and y-axis and six anatomical marker movements (mm) in non-dominant, eyes-closed condition of unipedal stance test (UPST) best trial† between footballers and sedentary students.

	Median (IQR)		
	Football (n=20)	Sedentary (n=20)	*p*-value
Complete UPST			
Time (s)	45.0 (21.2)	29.4 (29.2)	0.044*
CoP x-axis range (mm)	45.4 (19.0)	41.3 (20.2)	0.646
CoP y-axis range (mm)	41.4 (6.4)	42.1 (12.9)	0.626
Distance travelled in first 5-s of UPST			
1st metatarsal (mm)	100.6 (73.4)	103.5 (65.9)	0.534
5th metatarsal (mm)	73.6 (40.1)	72.1 (54.6)	0.787
Ankle (mm)	64.5 (44.8)	61.5 (30.4)	0.607
Heel (mm)	43.4 (26.3)	55.8 (18.7)	0.162
Knee (mm)	146.1 (107.7)	142.2 (87.5)	0.978
Greater trochanter (mm)	118.0 (68.7)	134.4 (54.6)	0.555

†Best trial is defined as the longest time taken and with the shortest CoP distance travelled during UPST.
IQR, inter-quartile range.

Table 4: Correlation between time for complete UPST time (s) and lower limb maker distances (mm) during the dynamic phase (n=36).

Lower limb sites	Time	
	r	*p*-value
First metatarsal	-0.50	0.001
Fifth metatarsal	-0.46	0.003
Ankle	-0.38	0.012
Knee	-0.44	0.004
Heel	-0.55	<0.001
Greater trochanter	-0.61	<0.001

Table 5: Multivariable linear regression model to predict the influence of greater trochanter, knee and ankle on time taken to complete unipedal stance test in collegiate footballers (n=18).

Variables	β	B	SE	*p*	95% CI
Greater trochanter	-0.77	-0.22	0.06	0.002	-0.35, -0.09
Knee	-0.09	-0.02	0.04	0.644	-0.09, 0.06
Ankle	-0.06	-0.02	0.05	0.685	-0.13, 0.09
Constant	-	69.1	5.2	<0.001	-

β, *standardized coefficient; B, unstandardized coefficient; SE, coefficients standard error; CI, confidence interval*

posture and neuromuscular control [9]. Thus, engaging in recreational sports, may it be football or other sports, may have added advantage to balance skills even in young adults. This may be a good foundation where balance is enhanced and hopefully maintained to prevent falls and fractures in later life.

The better balance performance in footballers compared to sedentary students were apparent in the sway observed, which was measured as the CoP distance travelled. Regardless if it was during the early dynamic phase (first 5-s) or static phase (after 5 s), footballers had less sway and were able to keep their movement to a minimal throughout the UPST on their non-dominant leg with eyes closed. The CoP distance was significantly shorter despite the range of CoP values in both the x- and y-axis being similar in footballers and sedentary students. Thompson and colleagues found out that footballers have better control in their anterior-posterior and mediolateral stability than the non-athletes [23]. This finding is in line with our outcome of this study.

Lastly, footballer have their hips to thank for keeping their UPST balance time longer compared to sedentary students. Their greater trochanter movement could significantly predict up to 77% of the time to complete the UPST while using their non-dominant leg and without visual input. This means participation in recreational football may indirectly provide players with better balance and postural control from the hip that may be missing in more sedentary people. As the hip is a known controller of postural stability [24], having specific and targeted exercises to improve hip control alone for balance may be difficult. Yet, participating in football games up to three times a week could train our balance intrinsically while having fun. Krustrup and colleagues mentioned that the training component of football were able to improve the postural balance of the players [25].

5. Conclusion

This is a cross-sectional study and our results are inferential at best. Our study sample size could have been improved as well to obtain a better normal distribution of data. However, non-parametric statistical analyses are equally powerful to parametric tests when used appropriately [26].

To improve on our balance protocol, we suggest future studies consider incorporating a warm-up component before performing UPST as it may improve overall balance [27]. With regards to the maximum time of 45 s, this may be suitable for young adults but researchers may consider different times for different type of population or even consider not having a time limit to include the possibility of muscle fatigue in different types of population.

The strength of our study highlights the need to understand muscle control towards balance in early dynamic phases that will contribute to the completion of the balance task at hand. Subsequently, use of advanced technology such as a force platform to assess CoP are ideal and provides a better understanding of the intricacies that are working to provide posture and stability in individuals of all ages.

Further studies are needed to support our current findings and intervention studies in sedentary young adults will be beneficial to inform in this area. More work can be done to understand the contribution of lower limb muscles and activation especially in the more vulnerable groups like older adults and people with osteoporosis to prevent falls and fractures.

Conflict of Interest

There is no conflict of interest reported between the authors.

Acknowledgment

This project was partly funded by USM Short Term Grant (304/PPSK/61312135).

References

[1] J. Winearls, "Posture: its function in efficient use of the human organism as a total concept" Agressologie, 13, 99-101, 1972.

[2] R. Lepres, A. X. Bigard, J. P. Diard, J. F. Gouteyron, C. Y. Guezennec, "Posture control after prolonged exercise" Eur J Appl Physiol Occup Physiol.. 76(1), 55-61, 1997. https://doi.org/10.1007/s004210050212

[3] P. G. Gerbino, E. D. Griffin, D. Zurakowski, "Comparison of standing balance between female collegiate dancers and soccer players" Gait Posture, 26(4), 501-507, 2007. doi:10.1016/j.gaitpost.2006.11.205

[4] D. A. Winter, A. E. Patla, J. S. Frank, "Assessment of balance control in humans". Med Prog Technol, 16(1-2), 31-51, 1990.

[5] T. Paillard, F. Noe, "Effect of expertise and visual contribution on postural control in soccer" Scand J Med Sci Sports, 16(5), 345-348, 2006. DOI: 10.1111/j.1600-0838.2005.00502.x

[6] J. Jerosch, L. Thorwesten, H. Bork, M. Bischof, "Is prophylactic bracing of the ankle cost effective?" Orthopedics, 19(5), 405-414, 1996. https://doi.org/10.3928/0147-7447-19960501-10

[7] B. L. Riemann, R. Schmitz, "The relationship between various modes of single leg postural control assessment" Int. J. Sports Phys. Ther., 7(3), 257-266, 2012.

[8] B .A. Springer, R. Marin, T. Cyhan, H. Roberts, N. W. Gill, "Normative values for the unipedal stance test with eyes open and closed" J Geriatr Phys Ther, 30(1), 8-15, 2007

[9] M. D. Jakobsen, E. Sundstrup, P. Krustrup, P. Aagaard, "The effect of recreational soccer training and running on postural balance in untrained men" Eur J Appl Physiol, 111(3), 521-530, 2011.DOI 10.1007/s00421-010-1669-2

[10] E. S. Chumanov, C. D. Remy, D. G. Thelen, D. G, "Computational techniques for using insole pressure sensors to analyse three-dimensional joint kinetics" Comput Method Biomec, 13(5), 505-514, 2010. DOI:10.1080/10255840903350148

[11] B. Pueo, J. M. Jimenez-Olmedo, "Application of motion capture technology for sport performance analysis" Retos, 32, 241-247, 2017.

[12] S. L. Colyer, M. Evans, D. P. Cosker, A. I. Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system" Sports Med-Open, 4(1), 24, 2018. https://doi.org/10.1186/s40798-018-0139-y

[13] E. Jonsson, A. Seiger, H. Hirschfeld, "One-leg stance in healthy young and elderly adults: a measure of postural steadiness?" Clin Biomech 19, 21(3), 688-694, 2004. doi:10.1016/j.clinbiomech.2004.04.002

[14] T. Muehlbauer, C. Mettler, R. Roth, U. Granacher, "One-leg standing performance and muscle activity: are there limb differences?" J Appl Biomech, 30(3), 407-414, 2014. https://doi.org/10.1123/jab.2013-0230

[15] N. Takeshima, N. L. Rogers, M. E. Rogers, M. M. Islam, D. Koizumi, S. Lee, "Functional fitness gain varies in older adults depending on exercise mode" Med Sci Sports Exerc, 39(11), 2036-2043, 2007. DOI: 10.1249/mss.0b013e31814844b7

[16] J. A. Kim, O. B. Lim, C. H, Yi, "Difference in static and dynamic stability between flexible flatfeet and neutral feet" Gait posture, 41(2), 546-550, 2015. https://doi.org/10.1016/j.gaitpost.2014.12.012

[17] A. Karlsson, G. Frykberg, "Correlations between force plate measures for assessment of balance" Clin Biomech, 15(5), 365-369, 2000. doi:10.1016/s0268-0033(99)00096-0

[18] A. Mansfield, E. L. Inness, "Force plate assessment of quiet standing balance control: Perspectives on clinical application within stroke rehabilitation" Rehabilitation Process and Outcome, 4, 7-5, 2015. doi:10.4137/RPO.S20363.

[19] T. Paillard, F. Noe, T. Riviere, V. Marion, R. Montoya, P. Dupui, "Postural performance and strategy in the unipedal stance of soccer players at different levels of competition" J. Athl. Train., 41(2), 172-176, 2006.

[20] R. Barone, F. Macaluso, M. Traina, V. Leonardi, F. Farina, V. Felice, "Soccer players have a better standing balance in nondominant one-legged stance" Open Access J Sports Med, 2, 1-6, 2011. doi:10.2147/OAJSM.S12593

[21] C. D. Davlin, "Dynamic balance in high level athletes" Percept Mot Skills, 98(3_suppl), 1171-1176, 2004.

[22] L. Bringoux, L. Marin, V. Nougier, P. A. Barraud, C. Raphel, "Effects of gymnastics expertise on the perception of body orientation in the pitch dimension" J Vestib Res, 10(6), 251-258, 2000.

[23] L. Thompson, M. Badache, S. Cale, L. Behera, N. Zhang, "Balance performance as observed by center-of-pressure parameter characteristics in male soccer athletes and non-athletes" Sports, 5(4), 86, 2017. doi:10.3390/sports5040086

[24] L. Allet, H. Kim, J. Ashton-Miller, T. De Mott, J. K. Richardson, "Frontal plane hip and ankle sensorimotor function, not age, predicts unipedal stance time" Muscle Nerve, 45(4), 578-585, 2012. DOI 10.1002/mus.22325

[25] P. Krustrup, P. Aagaard, L. Nybo, J. Petersen, M. Mohr, J. Bangsbo, "Recreational football as a health promoting activity: a topical review" Scand J Med Sci Sports, 20(Suppl. 1), 1-13, 2010. doi: 10.1111/j.1600-0838.2010.01108.x

[26] M. Hughes, S. M. Cooper, A. Nevill, "Analysis procedures for non-parametric data from performance analysis" Int J Perf Anal Spor., 2(1), 6–20, 2002. doi:10.1080/24748668.2002.11868257

[27] A. Daneshjoo, A. H. Mokhtar, N. Rahnama, A. Yusof, "The Effects of Comprehensive Warm-Up Programs on Proprioception, Static and Dynamic Balance on Male Soccer Player". PLoS ONE, 7(12): e51568, (2012) doi:10.1371/journal.pone.0051568

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 08-12 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

ANFIS-Based Climate Controller for Computerized Greenhouse System

Hassan Oubehar[1,*], Abdelouahed Selmani[1], Abdelali Ed-Dahhak[2], Abdeslam Lachhab[2], Moulay El Hassane Archidi[1], Benachir Bouchikhi[1]

[1]*Sensors Electronic & Instrumentation Team, Physic Department, Moulay Ismaïl University, Faculty of Sciences, Meknes, Morocco*

[2]*Modeling, Systems Control and Telecommunications Team, Department of Electrical Engineering, Moulay Ismaïl University, High School of Technology, Meknes, Morocco*

ARTICLE INFO

Article history:
Received: 07 October, 2019
Accepted: 12 December, 2019
Online: 15 January, 2020

Keywords:
Greenhouse
Climate control
Temperature
Relative humidity
ANFIS controller

ABSTRACT

The greenhouse climate system is very hard to manipulate because the variables involved are closely correlated. This study aims to enhance the regulation performance of greenhouse climate system based on adaptive neural-fuzzy inference system (ANFIS). The ANFIS is a hybrid technique that incorporates the fuzzy logic theory and artificial neural network algorithms. The employed control system has the ability to stabilize the climatic variables inside greenhouse system at required levels for crops development. The datasets utilized to train ANFIS model was obtained by implementing a fuzzy logic controller under MATLAB Simulink environment. The feasibility, reliability and robustness of the adopted strategy have been experimentally studied. The experiments conducted in real-time show that the proposed technique provides good tracking, short response time and high robustness with regard to outside perturbations and non-linear behaviors of greenhouse systems.

1. Introduction

During recent decades, computerized greenhouses become an important alternative for crops production to meet the food requirement that is increasing continuously due to the dramatic demographic growth. The quantity and quality improvements of fruits and vegetables need to maintain adequate environmental conditions under greenhouse. For that, growers have to control the main greenhouse climatic parameters that are important for plants growth [1] [2]. Nevertheless, the regulation of internal greenhouse atmosphere is a complicated process. Indeed, the different involved variables interact in a strong way. Moreover, the climate under greenhouse relies not only on the effect of external environment but also on the actions that are produced by the elements of the system through the climate control hardware [3]. During last years, various papers have investigated the smart control technologies of greenhouse production climates like PI-intelligent technique [4], fuzzy-based control systems [5-7], second-order sliding mode [8], nonlinear adaptive PID control [9] and model predictive control [10]…

*Corresponding Author: Hassan Oubehar, hassan.oubehar@gmail.com

Controllers based on conventional approaches are broadly implemented and they allow successful control for linear and time-invariant systems. Nevertheless, greenhouse is considered as complex process. In fact, it presents large nonlinear features, its parameters vary in time and also it has big reliance to the ambient environmental variables [11]. In response, the techniques of artificial intelligence (AI) were introduced in the literature to solve such issues. The main theories belonging to AI are fuzzy logic systems (FLS) and artificial neural networks (ANN). FL-based controller has many benefits for instance: the ability to reason like human being and to be easily designed based on sample dataset, so the exact mathematical model is not demanded. In addition, it has the characteristic to be adapted through changing the generated rules (there is no need to redevelopment) [12]. For the second part, ANN algorithms have great fastness and enormous learning capacity by means of a sample database and to establish nonlinear relationships between different dependent and independent variables [13]. To make use of the reasoning potential of FLS and powerful learning capacities of ANN, ANFIS algorithm has emerged for addressing the challenge of controlling complex, dynamic and non-linear systems. In this study, an ANFIS-based

control system was developed in order to promote an appropriate microclimate inside greenhouse system.

2. Schematic view of the hardware system

This study is conducted based on an automated experimental greenhouse as depicted in Figure 1. The considered prototype was home developed and installed at the Laboratory of Electronic, Automatic and Biotechnology that is a part of the Faculty of Sciences of Meknes in Morocco. A set of sensors are employed to ensure the measurements of various environmental variables involved in greenhouse system. LM35DZ sensors are utilized for measuring the outdoor and indoor temperature. The measurements of inside and outside relative humidity are performed using HIH-4000-001 sensors. To achieve control purposes, the experimental setup was equipped with an electrical heater and a variable-speed ventilator. All components are connected to an acquisition and control software units by means of a personal computer and data acquisition card NI 6024E [14-16].

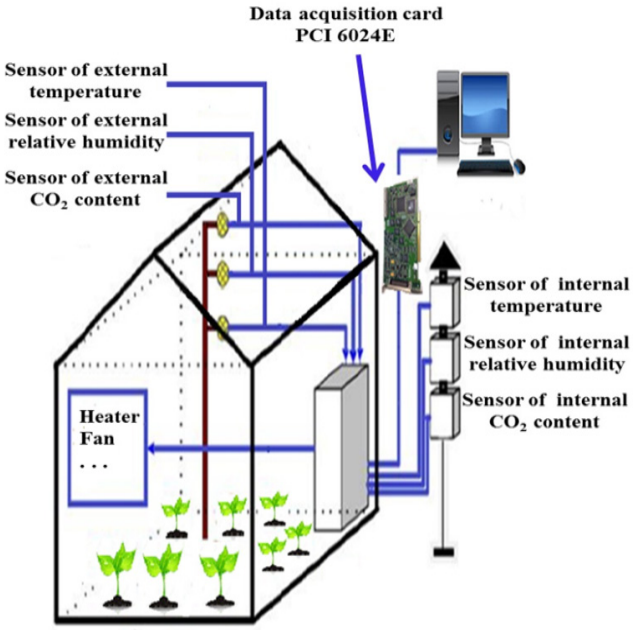


Figure 1: Experimental greenhouse system

3. ANFIS-based Control Design

ANFIS is an intelligent technique largely used by research communities for system modeling and control [17-19]. In fact, ANFIS is an efficient integration of fuzzy logic concepts and ANN algorithms. ANFIS approach are useful to build smart control technics that holds the ability of FLS to process the knowledge provided by human expert so as to imitate his actions by use of ANN learning capability [20]. ANFIS implements Sugeno-type fyzzy rules that are automatically tuned based on ANN algorithms. The fundamental principle of neuro-fuzzy systems is to create fuzzy-based model employing ANN methods where the fuzzy model parameters stand for the weights of the neural network structure [21].

Figure 2 shows the standard ANFIS architecture, where a circle denotes a fixed node, while a square represents an adaptive node. It's composed essentially of five layers (fuzzification, production, normalization, defuzzification and output layer). For simplicity, two inputs and one output are considered.

In the input layer, real values are transferred into linguistic variables. Each node is associated with a node function. The membership degree of linguistic variables is generated according to the following equation:

$$O_i^1 = \mu_{Ai}(x) \quad \text{for} \quad i = 1,2 \tag{1}$$

Here, x; Ai and μ point out input variable, fuzzy set and membership function of the i-th node respectively. If μ is generalized bell type, it is determined by three parameters {a,b,c}:

$$\text{bell}(x;a,b,c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \tag{2}$$

The second layer is called as product layer. Each node's output is obtained by multiplying incoming signals to it with each other using the following formula:

$$O_i^2 = w_i = \mu_{Ai}(x)\mu_{Bi}(y) \quad \text{for} \quad i = 1,2 \tag{3}$$

In the third layer, each node is a fixed one known as N. The i-th node computes the normalized firing strength or normalized weight by dividing the i-th firing strength with the sum of all rules firing strengths:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad \text{for} \quad i = 1,2 \tag{4}$$

In the fourth layer all nodes are adaptive ones where the node function is given by:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \quad \text{for} \quad i = 1,2 \tag{5}$$

Where $\bar{w}$ is a normalized firing strength from the previous layer and (pi, qi, ri) is the consequent parameter set of the node.

The single fixed node in the fifth layer calculates the overall output by the summation of all incoming signals using the following expression:

$$O_i^5 = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad \text{for} \quad i = 1,2 \tag{)}$$

The procedure of ANFIS design is demonstrated in Figure 3, which comprises essentially five steps. At first, the predefined dataset has been loaded into ANFIS-Editor. The following stage consists of converting the learning data into fuzzy data through a suitable membership function. Afterwards, the corresponding fuzzy rules are established based on fuzzification process. During the fourth stage, parameters of the generated model are updated

using the selected learning algorithm. At last, the defuzzification mechanism transfers the fuzzy output to real values and the constructed model is saved.

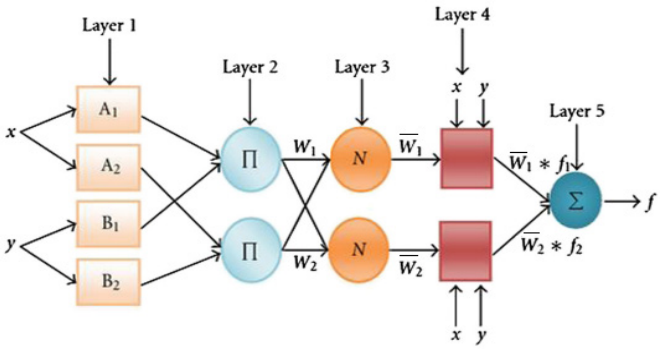


Figure 2: Basic ANFIS structure

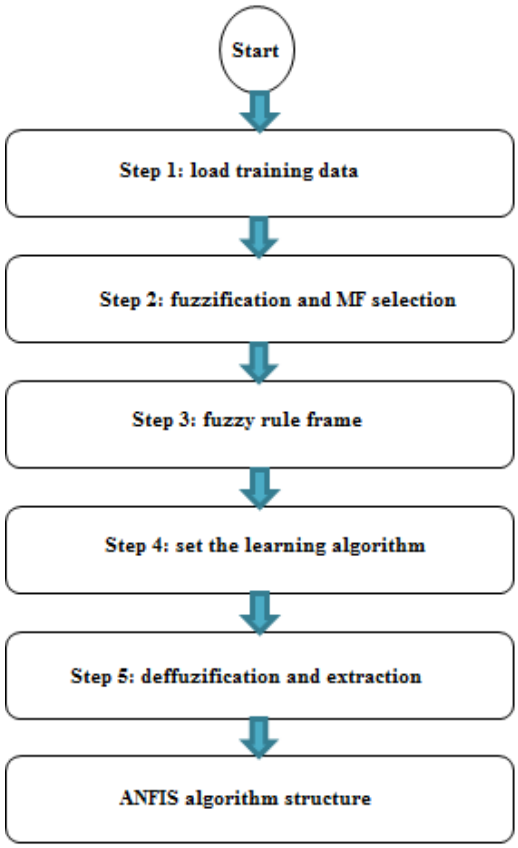


Figure 3: ANFIS flowchart

3.1. Greenhouse Climate ANFIS-based Control

The ANFIS controller developed in this works uses two inputs temperature tracking error e(T) and humidity tracking error e(RH) and a single output (voltage). A FL controller was simulated in MATLAB Simulink to provide pairs of inputs-output data. The collected database was recorded so as to be used later for training the adopted control scheme. Figure 4 illustrates the implemented control methodology. The considered control structure was applied to bring the indoor temperature (Tin) and indoor relative humidity (RHin) to desired trajectories (Ts) and (RHs). In fact, the suggested control system drives control actions via the power circuit to reach optimal climate conditions.

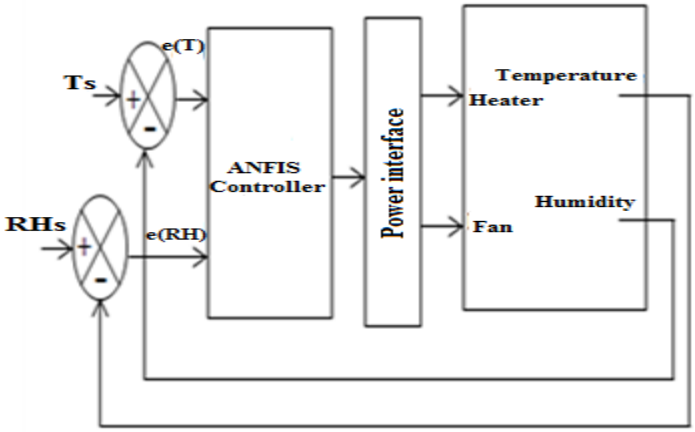


Figure 4: Schematic diagram of the proposed control process

To construct the ANFIS-based model five "generalized bell" membership functions was associated with every input as indicated in Figure 5 and Figure 6. The output variable was deffuzified using weighted average technique.

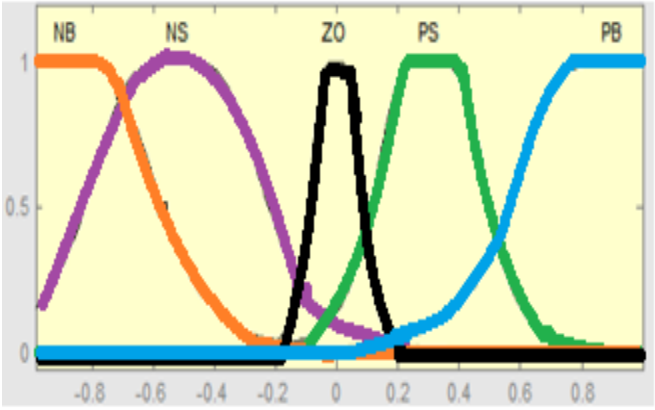


Figure 5: First input fuzzification

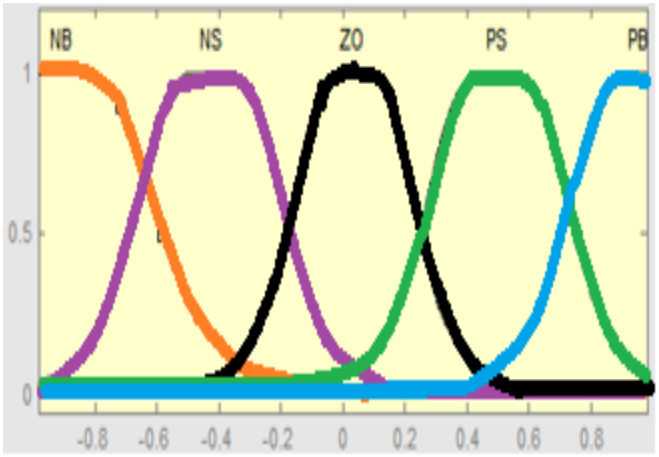


Figure 6: Second input fuzzification

The ANFIS model was created based on hybridization of two methods which consist of gradient technique to compute input parameters, and least square approach to calculate output function parameters. The architecture of the generated ANFIS controller is shown in Figure 7. The corresponding fuzzy rules are highlighted

in Figure 8. We note that the training data was normalized. Hence, the input values belong to the range [-1, 1] while the output voltage is between 0 and 1.

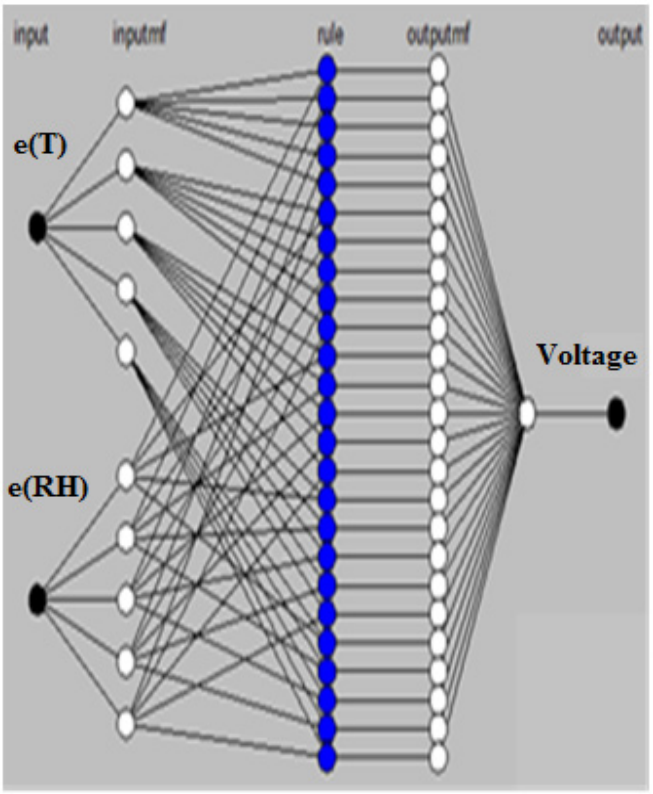


Figure 7: Structure the proposed control Model

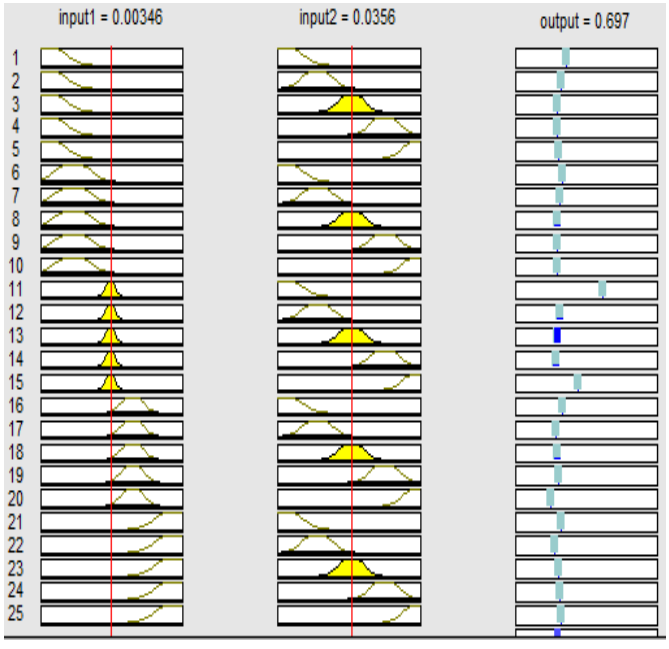


Figure 8: Fuzzy rules viewer

4. Experimental Results

The main experiments and performed measurements in real time are displayed in Figure 9 and 10. As it observed the indoor temperature (Tin) and indoor relative humidity (HRin) follow closely the predefined set points (Ts) and (RHs). The inside temperature remains close to the set point regardless of external temperature variations (Text). This proves that the developed control strategy is robust relating to external perturbations. As for the inside relative humidity, it is kept at required levels after few minutes. It is clearly obvious that the designed control system allows good regulation performance thanks to its robustness regarding disturbances due to the outside weather and system uncertainties. In fact, it permits to realize an accurate and quick response.

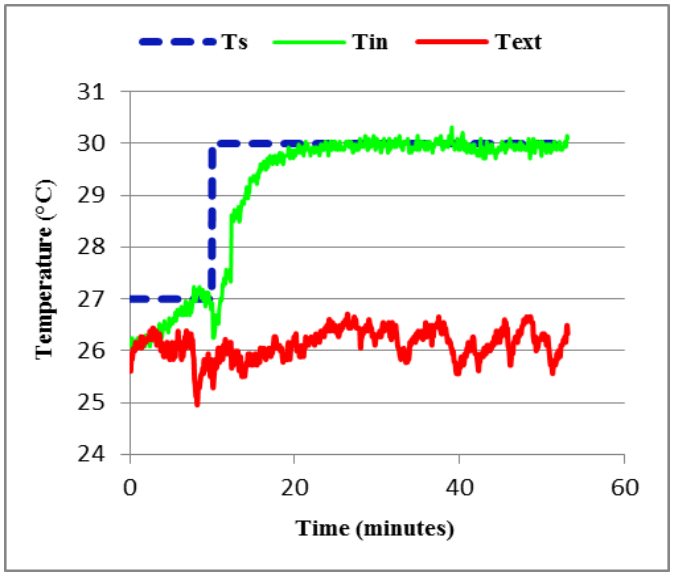
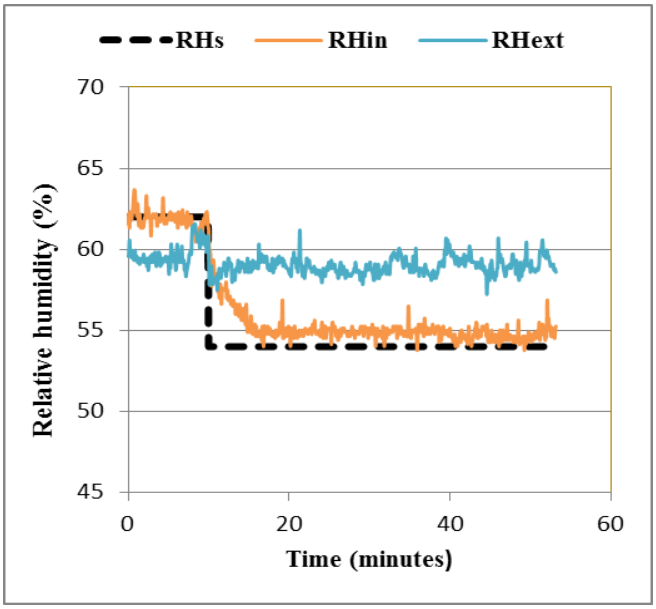


Figure 9: Evolution of internal and external temperatures



5. Conclusion

This work has developed an automation system on regulating different keys parameters of environmental greenhouse based on ANFIS algorithm. The proposed control strategy was successfully validated in real time. The experimental measurements demonstrate that the suggested technique is able to track the

reference trajectories quickly and without overshoot. It is clear that the used control method guarantees an accurate, robust and fast regulation of greenhouse indoor temperature and relative humidity to create optimal conditions for crops.

References

[1] H. Oubehar, A. Selmani, A. Ed-Dahhak, A. Lachhab, M.H. Archidi, B. Bouchikhi, "Design and real time implementation of ANFIS controller for greenhouse climate" in 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, 2018.

[2] M.A. Márquez-Vera, J.C. Ramos-Fernández, L.F. Cerecero-Natale, F. Lafont, J.F. Balmat and J.I. Esparza-Villanueva, "Temperature control in a MISO greenhouse by inverting its fuzzy model," Computers and Electronics in Agriculture, vol. 124, pp. 168–174, 2016.

[3] N. Bennis, J. Duplaix, G. Enea, M. Haloua and H. Youlal, "Greenhouse climate modelling and robust control," Computers and Electronics in Agriculture, vol. 61, pp. 96-107, 2008.

[4] Y. El afou, L. Belkoura, M. Outanoute, A. Rahali, A. Lachhab, B. Bouchikhi, "Feedback Techniques Using PID and PI Intelligent For Greenhouse Temperature Control" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,vol. 3, pp. 9779-9792, 2014.

[5] M. Guerbaoui, A. Ed-dahhak, Y. El Afou, A. Lachhab, L. Belkoura and B.Bouchikhi, "Implementation of Direct Fuzzy Controller in Greenhouse based on Labview" International Journal of Electrical and Electronics Engineering Studies, Vol.1 No.1, pp.1-13, September 2013.

[6] M.A. Márquez-Vera, J.C. Ramos-Fernández, L.F. Cerecero-Natale, F. Lafont, J.F. Balmat, and J.I. Esparza-Villanueva, "Temperature control in a MISO greenhouse by inverting its fuzzy model," Computers and Electronics in Agriculture, vol. 124, pp. 168–174, 2016.

[7] R. Ben Ali, S. Bouadila and A. Mami, "Development of a Fuzzy Logic Controller applied to an agricultural greenhouse experimentally validated," Applied Thermal Engineering, vol. 141, pp. 798–810, 2018.

[8] H. Oubehar, A. Ed-Dahhak, A. Selmani, M. Outanoute, A. Lachhab, M. Guerbaoui, M. H. Archidi and B. Bouchikhi, " High-Order Sliding Mode Control of Greenhouse Temperature" Indonesian Journal of Electrical Engineering and Computer Science, vol. 4, no. 3, pp. 548 -554, December 2016.

[9] S. Zeng , H. Hu, L. Xu and G. Li, "Nonlinear Adaptive PID Control for Greenhouse Environment Based on RBF Network, " Sensors, vol. 12, pp. 5328-5348, 2012.

[10] L. Chen, S. Du, Y. He, M. Liang, D. Xu, "Robust model predictive control for greenhouse temperature based on particle swarm optimization," Info Proc. Agri., in press.

[11] H. Hu, L. Xu, B. Zhu, and R. Wei, "A compatible control algorithm for greenhouse environment control based on MOCC strategy" Sensors, vol. 11, no. 3, pp. 3281-3302, 2011.

[12] L. Dambrosio, "Data-based fuzzy logic control technique applied to a wind system" Enregy Procedia, vol. 126, pp. 690-697, 2017.

[13] H.K. Ghritlahre, R.K. Prasad, "Application of ANN technique to predict the performance of solar collector systems," Renewable and Sustainable Energy Reviews, vol. 84, pp. 75–88, 2018.

[14] A. Ed-Ddahhak, A. Lachhab, L. Ezzine, and B. Bouchikhi, "Performance evaluation of a developing greenhouse climate control with a computer system," AMSE Journal Modelling, Measurement and Control, vol. 68, no. 1, pp. 53-64, 2007.

[15] A Ed-Dahhak, M. Guerbaoui, Y. El Afou, M. Outanoute, A. Lachhab, L. Belkoura, and B. Bouchikhi, "Implementation of fuzzy controller to reduce water irrigation in greenhouse using LabVIEW," International Journal of Engineering and Advanced Technology Studies, vol. 1, no. 2, pp. 12-22, 2013.

[16] M. Guerbaoui,Y. El Afou, A. Ed-Dahhak, Lachhab A, and B. Bouchikhi, "PC-Based automated drip irrigation system. International Journal of Engineering Science and Technology, " vol. 5, no. 1, pp. 221-225, 2013.

[17] C.E. Lachouri, M.M. Lafifi, K. Mansouri and A. Belmeguenai, "Adaptive Neuro-Fuzzy Inference Systems for Modeling Greenhouse Climate" International Journal of Advanced Computer Science and Applications, vol. 7, no. 1, pp.96-100, 2016.

[18] N. Mathur, I. Glesk and A. Buis, " Comparison of adaptive neuro-fuzzy inference system (ANFIS) and Gaussian processes for machine learning (GPML) algorithms for the prediction of skin temperature in lower limb prostheses, " Medical Engineering and Physics,vol. 38 pp. 1083–1089, 2016.

[19] D.M. Atia and H.T. El-madany, "Analysis and design of greenhouse temperature control using adaptive neuro-fuzzy inference system, Journal of Electrical Systems and Information Technology," vol. 4, pp. 34–48, 2017.

[20] M. Zare and M. Koch, "Groundwater level fluctuations simulation and prediction by ANFIS- and hybrid Wavelet-ANFIS/Fuzzy C-Means (FCM) clustering models: Application to the Miandarband plain," Journal of Hydro-environment Research , vol. 18, pp. 63–76, 2018.

[21] B. Haznedar and A. Kalinli, "Training ANFIS structure using simulated annealing algorithm for dynamic systems identification," Neurocomputing, vol. 302, pp. 66–74, 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 13-20 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Industrial Network for The Control and Supervision of The Acetic Acid Dispatch Process, and Its Influence on The Reduction of Chemical Contaminants for Operators

Omar Chamorro-Atalaya[1,*], Jaime Yataco-Yataco[2], Dora Arce-Santillan [1]

[1]*Faculty of Engineering and Management, Universidad Nacional Tecnológica de Lima Sur, Perú*

[2]*Grupo ODFJELL Terminals, Depósitos Químicos Mineros, Lima, Perú*

ARTICLE INFO

Article history:
Received: 11 September, 2019
Accepted: 09 December, 2019
Online: 15 January, 2020

Keywords:
Control
Supervision
SCADA
Distributed Control
Occupational Health and Safety
Programming
Acetic acid

ABSTRACT

This article develops the design and implementation of an industrial network for the control and supervision of the process of dispatching acetic acid, which aims to reduce the presence of this chemical in the environment, since the process is activated manually by the operators; what generates a lack of precision in several stages of the process, that brings in many occasions, spill of the acetic acid that is highly flammable and harmful for the health of the operators; For this reason, an industrial network controlled by a Siemens S7-1200 Programmable Logic Controller (PLC) was developed using the SIMATIC STEP 7 Software, and a Siemens S410 Distributed Control System (DCS), using the PCS7 V 8.1 Software; where the control logic was programmed and the distributed periphery was implemented. Once the industrial network is implemented, the result is a 19.44% decrease in the presence of acetic acid in the environment; this value is complying with what is established in the regulation of "Permissible Limit Values of Chemical Agents in the Work Environment".

1. Introduction

Currently, almost all operators are exposed to some kind of risk, especially if the work to be done is linked to the use of chemical inputs; In fact, the risks of contact with chemical inputs are considered the most harmful in human health. Therefore, industries are in need of safeguarding the health of their operators; through the use of technologies, capable of reducing or mitigating the presence of chemical contaminants. As part of the strategies there is remote control, capable of supervising and controlling industrial processes from distant points. [1]

In other words, occupational health and safety requirements for operators and the environment, makes current industries have secure systems that allow generating data records and interact in a comfortable way through graphic interfaces, these systems are called SCADA (Supervisory Control and Data Acquisition) or control systems and process supervision. [2]

In a study carried out in Sangolqui, Ecuador, it is obtained that the implementation of these remote control and supervision systems (SCADA) has an effective communication infrastructure, which allows a 61.66% reduction of the risks to the operators in your industrial processes [3]

Likewise, in the industrial network of control and supervision of the crude oil industry it has been determined that the functionality achieved with the automation of the LACT units (Lease Automatic Custody Transfer) or Automatic Measurement Units for Custody Transfer, is 93.8 %, these results were able to meet the needs of the process, complying with the parameters established by the ARCH (Agency for Control and Regulation of Hydrocarbons) and ensuring the integrity of the unit of its workers before operational errors. [4]

The automatic control of emptying and filling of oil, reduces the intervention of the operator by 12% (18 minutes), and generates an optimal control in the spill of the liquid which is of high risk for the operators. [5]

In this sense, the present work aims to determine the influence of the decrease of chemical contaminants for the operators, through the design and implementation of an industrial network that will control and supervise the process of acetic acid dispatch.

The present work is structured as follows. Section 2 presents the methodology used in this research work; This methodology will have an Experimental Design of Quasi Experimental type, with a quantitative approach, since there will be a comparison of before and after the results obtained in relation to the presence of acetic acid in the environment. Section 3 shows the results

*Omar Freddy Chamorro Atalaya. Jr. Los Damascos 986. Los Olivos, Lima, Perú, 968053089 & omar_chamorro1@hotmail.com

obtained after observing the technical reports. Section 4 presents the statistical analysis. Finally, in Section 5 the discussion of the results obtained is presented.

2. Methodology

The process followed for the development of the proposal is as follows:

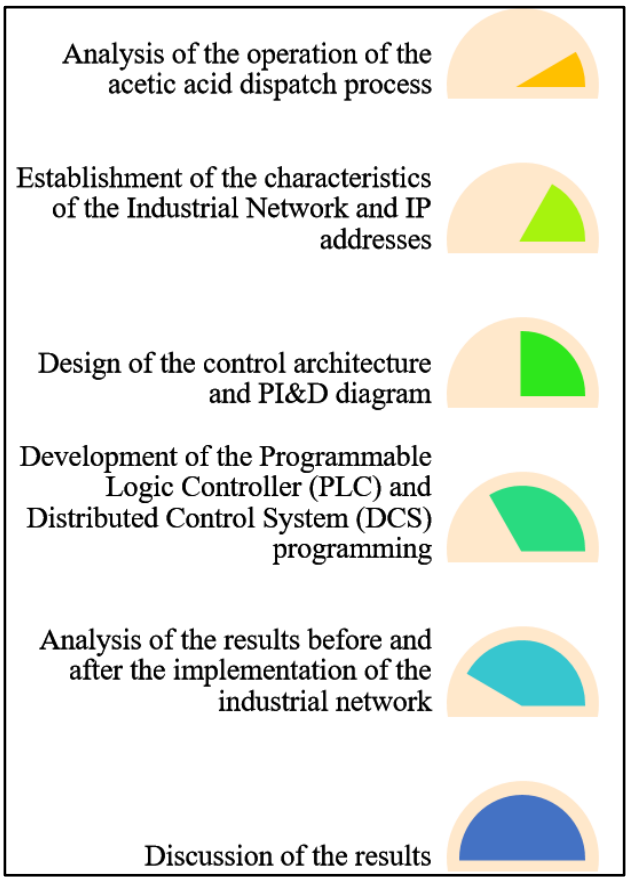


Figure 1: Development procedure

Once the procedure has been established, then the description of the system to be automated is carried out, with the purpose that said description establishes the context over which the guidelines for the development of the industrial network will be given, which will supervise and control the process of Chemicals dispatch.

According to the analysis performed, during the dispatch operation of trucks containing acetic acid, a series of equipment that are involved in the system are used, which require the intervention of the operators to perform the following activities:

- Valve maneuvers.
- Motor operation.
- Monitoring of constant connection of ground system against static loads.
- Registered operations.
- Permanent monitoring of possible overfill during the tanker truck dispatch process.

The steps to perform the dispatch operation are described in the following table:

Table 1: Steps to be performed in the Office Operation

Steps	Supervisor	Operator 1	Operator 2	Operator 3
Step 1: check permits office	■			
Step 2: truck income		■		
Step 3: take tare	■			
Step 4: pump ignition			■	
Step 5: valve opening		■		
Step 6: pump and valve monitoring		■	■	
Step 7: final weight monitoring	■			
Step 8: procedural valve closure		■		
Step 9: pump off			■	
Step 10: connection and disconnection				■

Source: self-made

As shown in the table above, 4 people are needed for the operation of the dispatch process; there must be constant communication between the supervisor and the operators for the control and registration of the offices. Dispatch trucks have a capacity of 30 t and have up to 5 compartments.

Taking as reference the described, next, we will proceed to describe the characteristics of the industrial network for the control and supervision of the process of clearance of acetic acid; which is comprised of:

- CCM Electrical Room
  - Communications board
  - PLC cabinet
  - Balance monitoring cabinet
  - Network analyzer cabinet
- Server Room

- Computer Technology Cabinet
- Industrial Instrumentation Cabinet

- Balance control room
  - Customer stations
  - Engineering station

The control architecture of the Industrial Network fulfills 3 important functions:

- Record historical data.
- Supervise the plant, safety interlocks and protection.
- Communicate with the existing DQM-1 system. (Future)

To guarantee these functions, the system has the following characteristics:

- The industrial network data server is located in the server room and is connected through a fiber optic network with the DQM-1 control network. (Future)
- The PLC was installed in the Industrial Instrumentation Cabinet, inside the DQM-2 electrical room.
- The "control switch" communication switch will allow the exchange of information between the PLC and the data server.
- All remote start / stop signals from each PLC will be using the control logic and sent to the CCM via Profibus DP communication.

Communication between the main Controller and the Supervision System is via Ethernet. For this, 100/1000 Mbps high speed industrial switches were considered.

For the automated industrial network, the use of:

- A PC called the Operating Station; it is located in the control room. This PC performs the operation, supervision and control in DQM-2. The monitoring software considered for this PC is of the "runtime" type.
- A PC called an Engineering Station is installed in the control room, intended for operation and maintenance work. The monitoring software considered for this PC is of type "Runtime & Development".
- Implementation of the card at the S7-300 level, projection of energy management through an industrial network branch in Mod Bus protocol.
- A Data Server to manage all the information that comes from the Control System. The server must include TCP / IP Ethernet connectivity and two next-generation network cards have been considered.

The communication network of the control system of the DQM-2 plant has the ability to communicate through a communication platform based on Ethernet TCP / IP and allows your data server to communicate with the server of the DQM-1 plant, to allow data from the new control system and the existing control system to be monitored from both plants. This link will be through fiber optics.

For the energy stage and grounding system the following was considered:

- Each cabinet has a copper bar to concentrate the connection of the Profibus DP fieldbus cable shields, analog 4–20 mA signal. And supply 24 VDC. An earth well was connected, which is connected to a group of earth wells near the Control Room or Electrical Room.

All analog signal instruments (transmitters) have Profibus DP communication. The discrete signal instruments and control signal switch contacts (switches, position switch, etc.) are of the dry contact type for an operating voltage of 110 VAC / 60Hz.

Following the development of the industrial network for the control and supervision of the acetic acid dispatch process, the following Table shows the established IP addresses:

Table 2: IP addresses

	Currently		**IP Assignment**	
	IP Address	***Terminal Name***	***Equipment Bus***	***Plant Bus***
DQM-02	Industrial Server	OSS1A	172.168.1.100	172.168.0.100
	Engineering Station	INSTRUDQM2	172.168.1.21	172.168.0.21
	CPU-PLC			172.168.0.10
	HMI Balance Room			172.168.0.50
	HMI GARITA			172.168.0.52
	Balance 10		172.168.1.110	
	Balance 11		172.168.1.111	
	Balance 12		172.168.1.112	
	XSCALANCE01	SW01-TB-SVR	172.168.1.181	
	XSCALANCE02	SW02-PB-SRV		172.168.0.182
	XSCALANCE03	SW03-TB-CCM	172.168.1.183	

Source: self-made

After the description of the characteristics of the industrial network for the control and supervision of the process of acetic acid dispatch in the previous point, then I will proceed to represent the control architecture, for which I highlight the components and the type of communication protocol used in the process of data transfer, between the CCM Electrical Room, server room, balance control room and field elements.

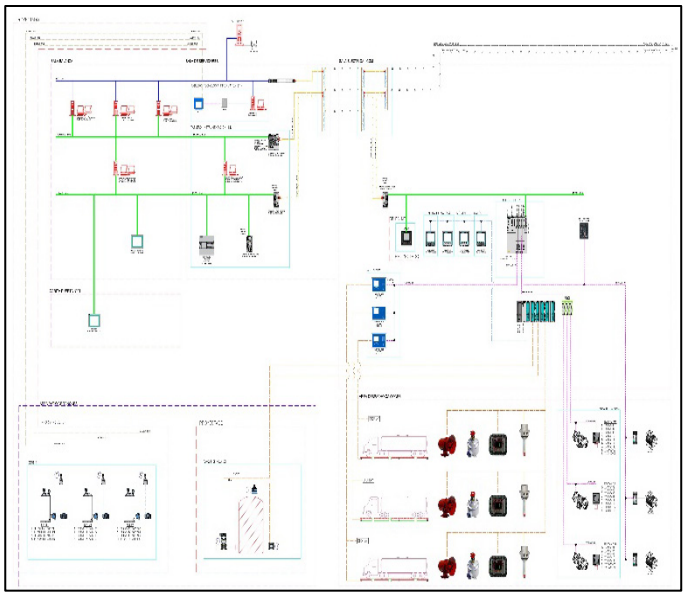


Figure 2: Industrial network control architecture

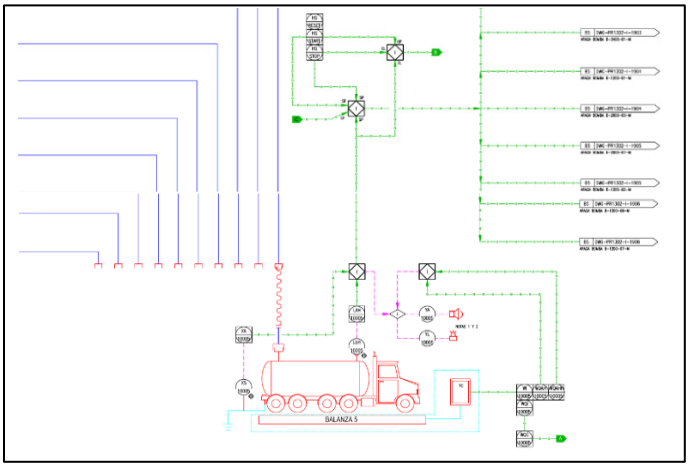


Figure 3: PI&D diagram

Next we proceed to develop the programming of the Siemens S7-1200 Programmable Logic Controller using the SIMATIC STEP 7 Software, and then perform the programming of the Distributed Control System (DCS SIEMENS S410), through the PCS7 V8.1 Software, where I configure the control logic.

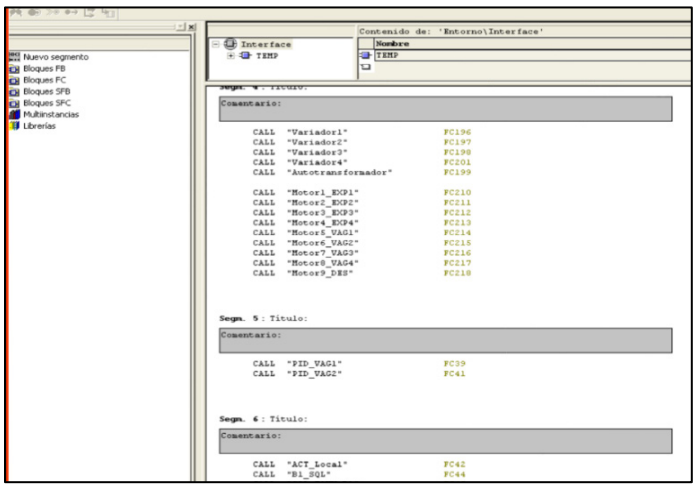


Figure 4: OB1 program execution block

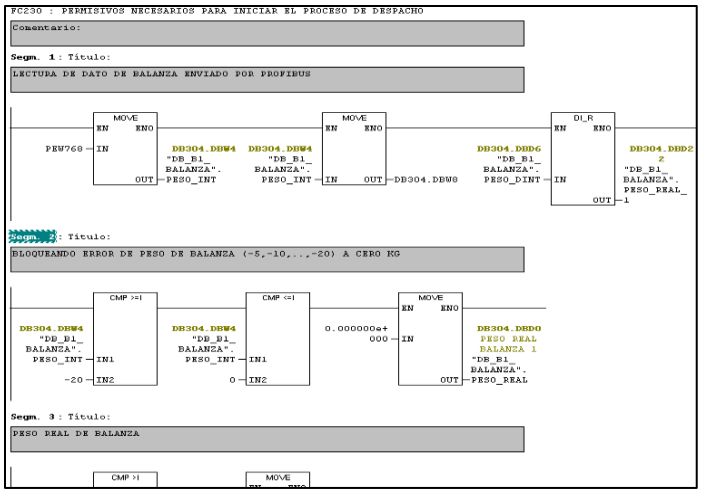


Figure 5. Ladder language programming

The monitoring system developed has areas for specific use and is always available to the user. The primary purpose of these areas is that the user can be immediately warned of any problem that occurs in the process and can also quickly go to review any screen of the process that is necessary. Likewise, the criterion that access to control and supervision is restricted and that the operation of the equipment is only carried out by authorized users was used.

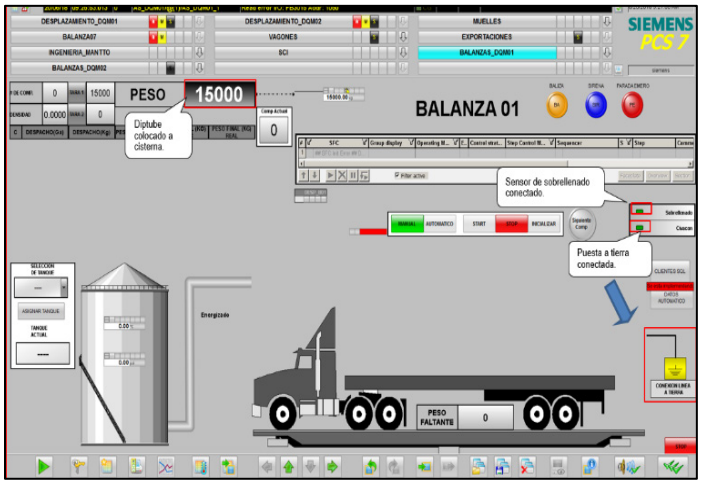


Figure 6: Grounding connection and overfill sensor connection

Figure 7: Generation of the operation code

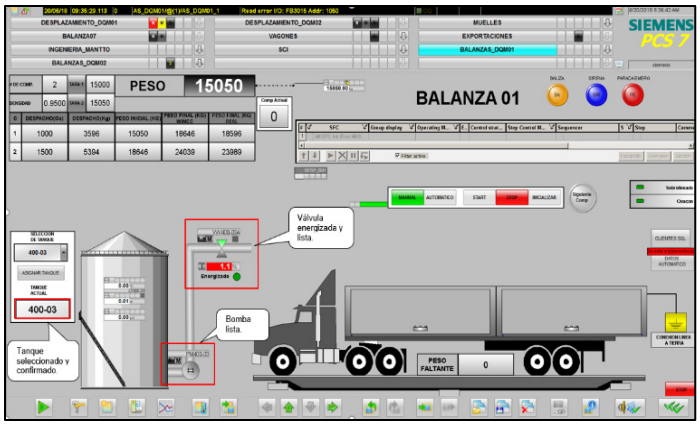


Figure 8: Enabled pump and energized valve for office 1

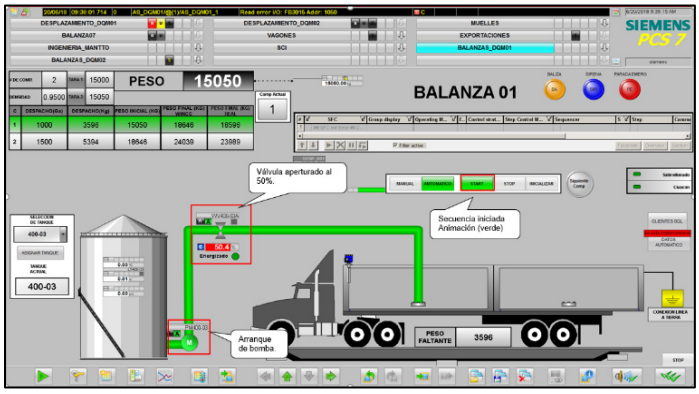


Figure 9: Starting the dispatch pump and opening the 50% electric valve

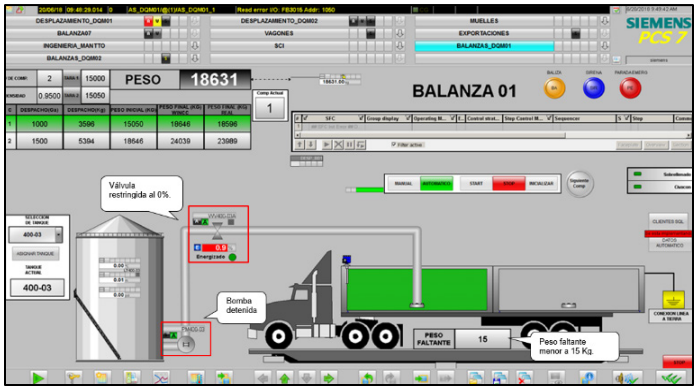


Figure 10. Electric valve completely closed, for a lack weight less than 15 kg - compartment 1

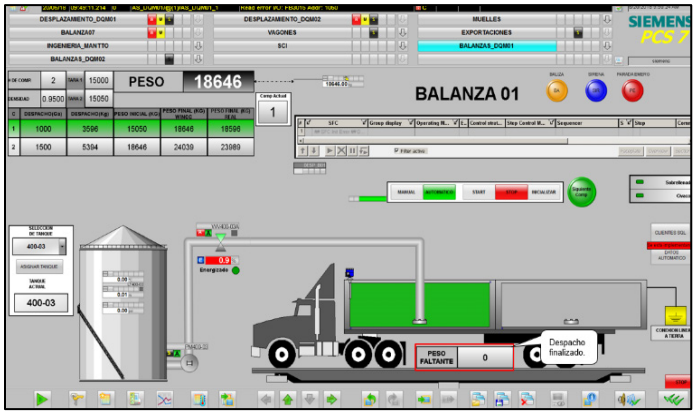


Figure 11. Finished office - compartment 1

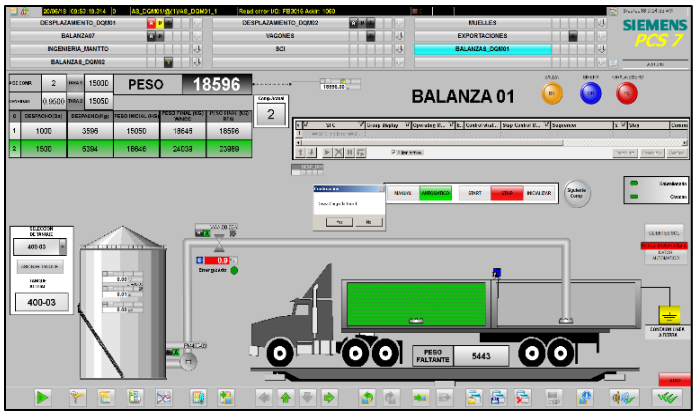


Figure 12: Compartment dispatch configuration 2

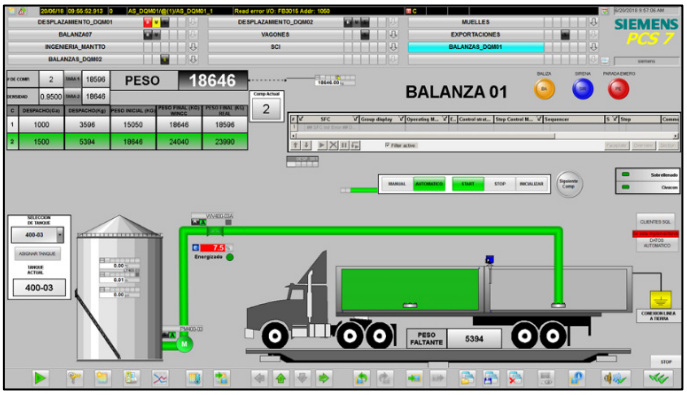


Figure 13: Dispatch of compartment 2

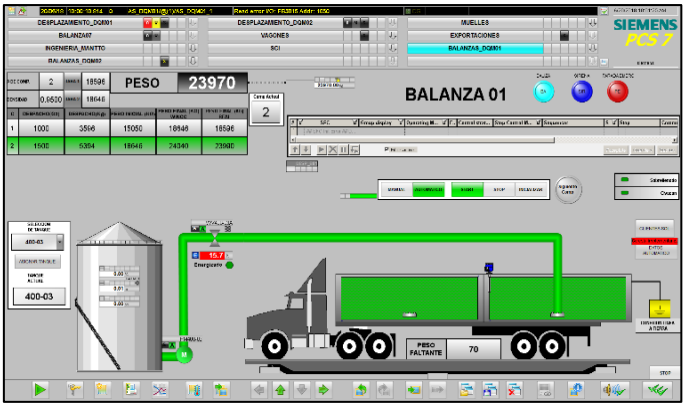


Figure 14. Electrical valve restricted to 15%, for a lack weight less than 70 kg. - compartment 2

Figure 15: finished office - removal from the tank

3. Results

In order to determine the influence on the decrease of chemical contaminants for operators; It was established that the research design is Experimental of Quasi-Experimental type, since there will be a comparison of before (pre-test) and after (post-test), the presence of acetic acid in the environment.

Our choice is based on the fact that, with the quasi-experimental design, it is intended to study the impact or influence of treatments and / or processes of change, in situations where the subjects or units of observation have not been assigned according to a criterion random. [6]

Likewise, the population is constituted by the Control and Supervision System of the acetic acid dispatch process, which is constituted by the CCM Electrical Room, the Server Room and the Balance Control Room; since, the population is defined as the totality of the phenomenon to be studied where population units have a common characteristic of which is studied and gives rise to the research data. [7]

The sample is equal to the population, since; if the population is less than fifty, the population is equal to the sample or the unit of analysis. [8] The type of sampling is non-probabilistic of the type "Deliberate, critical or by trial", which is an effective technique in situations where there is only a restricted number of units of analysis with similar qualities; this type of sampling generates that the results obtained are highly accurate with a minimum margin of error. [9]

In the present investigation 63 measurements have been made, which were taken in an interval of 1 hour, in a period of one week, that is to say, 9 measurements were made, both before (pre-test) and after (post- test) of the presence of acetic acid in the environment; these results were processed with the statistical software SPSS version 24. Given the above, the data processing is shown below.

	MANUAL_CONTAMINATION_LEVEL	AUTOMATIC_CONTAMINATION_LEVEL
1	25,0	20,0
2	26,5	20,0
3	26,5	20,0
4	26,5	21,5
5	26,0	21,0
6	26,5	21,5
7	26,0	21,0
8	27,0	22,0
9	27,0	22,0
10	27,0	22,0
11	27,0	22,0
12	27,0	22,0
13	27,0	22,0
14	27,0	22,0
15	27,0	22,0
16	25,0	20,0
17	25,0	20,0
18	26,0	21,0
19	26,0	21,0
20	26,0	21,0
21	26,0	21,0

Figure 16. Data processing

Likewise, in order to determine the influence of the decrease of chemical contaminants for the operators, through the design and implementation of an industrial network that will control and supervise the process of dispatching acetic acid; the statistical tool "Arithmetic Media" was used; The following Table shows the result of the measurements in relation to the presence of acetic acid in the environment.

It should be noted that according to the regulation of "Permissible Limit Values of Chemical Agents in the Work Environment" approved by the D.S. N° 015-2005-SA, by the Ministry of Health of the Peruvian state, the permissible limit value of acetic acid is 24.5 mg/m^3. [10]

Table 3. Average level of acetic acid pollution before and after the industrial network

	N	Minimum	Maximum	Half
Manual Contamination Level	63	25,0	27,5	26,206
Automatic Contamination Level	63	20,0	22,5	21,111
Valid N (per list)	63			

Source: SPSS

To determine the extent to which the industrial network influences the control and supervision of the acetic acid dispatch process in the reduction of chemical agents; we will use the following formula; which consists in calculating the average for each of the groups (manual and automatic); The difference between the two average observations will be the influence or impact of the present investigation.

$$\%\ de\ influence = \frac{AAI - ABI}{MAI}\ x\ 100 \quad \text{...(1)}$$

Where:

% influence = Percentage influence

AAI = Average after implementation

ABI = Average before implementation

According to the results in table 3, we replace the values in equation 1, obtaining the following:

$$\%\ de\ influence = \frac{21.111 - 26.206}{26.206}\ x\ 100$$

$$\%\ de\ influence = -19.44\%$$

According to the result, the percentage of influence has decreased by 19.44%; although this value seems small, it is very significant, since with this decrease, it is complying with what is established in the regulation of "Permissible Limit Values of Chemical Agents in the Work Environment".

Once the data analysis has been obtained and the influence result obtained, the normality test will be carried out in order to determine if the data is reliable.

The following levels of significance are considered: [11]

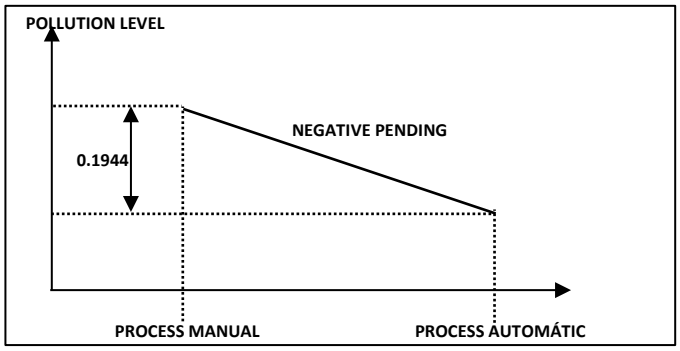


Figure 17. Representation of the decrease of acetic acid in the environment

- 0.05 for research projects
- 0.01 for quality research
- 0.10 to find markets and policies

So for the present investigation the level of significance (p) is 0.05 and implies that the researcher has a 95% reliability; below is the following statistical criteria established for this test:

- If the value p < 0.05, the data comes from a normal distribution
- If the value p >0.05, the data does not follow a normal distribution

The criterion to determine which values to use in the normality test is the following:

- Kolmogorov -Smirnov / large samples (> 30)
- Shapiro Wilk / small samples (<30)

In our case we will use the Shapiro Wilk test; the following table shows the result obtained:

Table 4: Normality test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistical	*GL*	*SIG.*	*Statistical*	*GL*	*SIG.*
Manual Contamination Level	,186	63	,000	,871	63	,000
Automatic Contamination Level	,229	63	,000	,814	63	,000

Source: SPSS

With the result of the normality test we can interpret that the values obtained are 95% reliable, that is to say that the data obtained have a normal distribution (True values).

Finally, we will determine the level of significance of the decrease result, using the T-Student test; this test is used when the following requirements are met:

- When it is possible to calculate the mean and standard deviation from the sample.
- The sample size must be less than 30.

For the development of this test we will initially calculate the degrees of freedom, which are the numbers of values of a sample that we can freely specify, after we already know something about that sample (n = 63), the calculation refers to the number of quantities of the sample, in this case they are the measurements made, minus the number of restrictions that bind the observations and the statistic, then the degree of freedom is as follows:

Degree of freedom: $n - 1$

Replacing in (2): $Degree\ of\ freedon = 62$

Once the degree of freedom is obtained, we will use the T-Student table to determine the critical value, with this we will statistically corroborate the influence of the industrial network variable on chemical pollutants.

Table 5: T-student

$n \setminus \alpha$	0,30	0,25	0,20	0,10	0,05
1	0,7265	1,0000	1,3764	3,0777	6,3137
2	0,6172	0,8165	1,0607	1,8856	2,9200
3	0,5844	0,7649	0,9785	1,6377	2,3534
4	0,5686	0,7407	0,9410	1,5332	2,1318
5	0,5594	0,7267	0,9195	1,4759	2,0150
6	0,5534	0,7176	0,9057	1,4398	1,9432
7	0,5491	0,7111	0,8960	1,4149	1,8946
8	0,5459	0,7064	0,8889	1,3968	1,8595
9	0,5435	0,7027	0,8834	1,3830	1,8331
10	0,5415	0,6998	0,8791	1,3722	1,8125
11	0,5399	0,6974	0,8755	1,3634	1,7959
12	0,5386	0,6955	0,8726	1,3562	1,7823
13	0,5375	0,6938	0,8702	1,3502	1,7709
14	0,5366	0,6924	0,8681	1,3450	1,7613
15	0,5357	0,6912	0,8662	1,3406	1,7531
16	0,5350	0,6901	0,8647	1,3368	1,7459
17	0,5344	0,6892	0,8633	1,3334	1,7396
18	0,5338	0,6884	0,8620	1,3304	1,7341
19	0,5333	0,6876	0,8610	1,3277	1,7291
20	0,5329	0,6870	0,8600	1,3253	1,7247
21	0,5325	0,6864	0,8591	1,3232	1,7207
22	0,5321	0,6858	0,8583	1,3212	1,7171
23	0,5317	0,6853	0,8575	1,3195	1,7139
24	0,5314	0,6848	0,8569	1,3178	1,7109
25	0,5312	0,6844	0,8562	1,3163	1,7081
26	0,5309	0,6840	0,8557	1,3150	1,7056
27	0,5306	0,6837	0,8551	1,3137	1,7033
28	0,5304	0,6834	0,8546	1,3125	1,7011
29	0,5302	0,6830	0,8542	1,3114	1,6991
30	0,5300	0,6828	0,8538	1,3104	1,6973
40	0,5286	0,6807	0,8507	1,3031	1,6839
80	0,5265	0,6776	0,8461	1,2922	1,6641
120	0,5258	0,6765	0,8446	1,2886	1,6576
∞	0,5244	0, 6745	0,8416	1,2816	1,6449

Source: (Campos, 2010)

The calculated critical value is 1.6706; with the critical value found, we will determine the significance value of the decrease in acetic acid in the environment, using the T-Student Test, of the SPSS statistical program.

Table 6: Test T - student

	95% trust interval for the difference		GL	Sig. Bilateral
	Lower	*Higher*		
Manual Contamination Level Automatic Contamination Level	4,9131	5,2774	62	,000

Source: SPSS

Next, we will analyze the results of the T-Student test (Bilateral Significance), and replace the result in the following expression.

The acceptance region is:

$$-\infty < t <= +1,6706 \quad \ldots(3)$$

$$t = 0,000$$

$$-\infty < 0,000 < 1,6706$$

With the result obtained, we can confirm that there is an influence of the industrial network variable on the variable polluting chemical agents, that is, the variables are related to each other. Also, since the significance value is less than 0.05, it means that there is a high level of significance in the decrease in acetic acid in the environment.

4. Discussion

In relation to the results of this research, the following discussions are held below:

Regarding the decrease of chemical contaminants for operators; in another study, it is pointed out that when implementing the SCADA system, it ensures that the operational conditions of the gas pipeline where natural gas is transported are controlled in real time, efficiently and safely. And in this way it contributes to the reduction of damage to facilities, workers, neighbors and the environment. [12]

Within the same approach in the investigation entitled Automation in the process of industrial painting in the drying stage, it is pointed out that when the control and supervision system was launched, many advantages were observed; Since, this mechanism generated an optimum precision in the drying stage, this was of utmost importance for the factory, since by emitting the toxic chemicals present in the paints in a smaller amount, thus reducing the risk of the health and safety of its personal. [13]

Another investigation with similar results, indicates that the automated system achieved the reduction of particles in the production of detergents, because the system was in charge of optimally controlling and supervising the entire process, generating that the staff has less contact with it (pumps, corrosive and flammable materials), thus achieving a significant increase in operator safety. In addition, it was demonstrated that the control system brings numerous operational, safety, production and monetary benefits. [14]

5. Conclusions

It is concluded that it was possible to determine the influence of the implementation of the industrial network for the control and supervision of the acetic acid dispatch process on the decrease of chemical contaminants for the operators, which is 19.44%; Although this value seems small, it is very significant, since with this decrease, it would be complying with what is established in the regulation of “Permissible Limit Values of Chemical Agents in the Work Environment”.

It is concluded that it was possible to design and implement an industrial network for the control and supervision of the acetic acid dispatch process, through the programming of the Siemens S7-1200 Programmable Logic Controller (PLC) through the SIMATIC STEP 7 Software, and a System of Distributed Control (DCS) Siemens S410, through the PCS7 V 8.1 Software, where the control logic was programmed and the distributed periphery for digital inputs and outputs was implemented to receive and emit signals from the field.

Finally, it is concluded that this project shows a high level of significance of 0.000, in relation to the decrease in acetic acid in the environments; also by means of the normality test we can interpret that the obtained values are 95% reliable, that is to say that the obtained data have a normal distribution; Therefore, the project contributes to occupational health and safety standards for operators and environmental protection.

Acknowledgment

The authors wish to acknowledge and thank the ODFJELL Terminals Group, Mining Chemical Deposits, Lima-Peru.

References

[1] Yataco, J. (2019). Design of an industrial network for the control and supervision of the chemical products dispatch process, to reduce the operator's exposure to harmful vapors, in the company Depósito Químicos Mineros S.A. ”. (Undergraduate thesis). Peruvian Wings University, Lima.

[2] Colomer, J., Meléndez, J., & Ayza, J. (2010). Introduction to the monitoring and expert supervision of processes. Methods and tools. Spanish Automatic Committee: Automated.

[3] Cueva, G., & Villa, A. (2019). Design and implementation of a SCADA system for the CEDIT industrial process simulation line, based on a chicken processing chain. (Undergraduate thesis). University of the armed forces, Sangolquí.

[4] Sandoval, C. (2017). Automation of LACT units for crude oil in the fields. University of the Armed Forces-ESPE, 1-6.

[5] Estévez, C., & Garcés, I. (2014). Design and simulation of a station to automate the process of emptying and filling with oil the modules of the tools of the Schlumberger company that operate at the bottom of the well during drilling. University of the Armed Forces, 1-8.

[6] Arnau, J. (1997). Experimental designs in schemes. Barcelona: University of Barcelona.

[7] Tamayo and Tamayo, M. (2003). The process of scientific research. Mexico: LIMUSA.

[8] Hernández, R., Fernández, C. B., & Pilar, M. d. (2003). Investigation methodology . Mexico: Mc Graw Hill.

[9] Landa, M., & Ramírez, M. (2018). Design of a student satisfaction questionnaire for a professional level course under the inverted learning model. Education Pages Magazine.

[10] Ministry of Health of the Peruvian state (2005). D.S. N ° 015-2005-SA. Perú

[11] Manterola, C., & Pineda, V. (2008). The value of "p" and the "statistical significance". Chilean Magazine, 60: 86-9.

[12] Padilla, I. (2015). Automation of a valve in a sectioning station and in a devil trap of gas pipelines. (Undergraduate thesis). National Polytechnic Institute, Quito.

[13] Ceballos, J., Hidalgo, S., & Salazar, J. (2014). Automation in the industrial painting process in the drying stage. (Undergraduate thesis). National Polytechnic Institute, Quito.

[14] Guiterrez, J. (2013). Design of an automated system for the production of disinfectants of a detergent plant. (Undergraduate thesis). Simon Bolivar University, Venezuela.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 21-33 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Storing and Transporting Hazardous Material, Logistics Strategies for Moroccan Companies

Abdellatif Soussi*[1,2], Dounia Bouchta[1], Roberto Sacile[2], Chiara Bersani[2], Hamid Seghiouer[3], Ahmed El Amarti[1], Rachid Abouettahir[3], Redouan El Khamlichi[1]

[1]*Laboratory of Materials and Interfacial Systems Faculty of Sciences, UAE Tétouan, 93000, Morocco*

[2]*Department on Informatics, Bioengineering, Robotics and Systems Engineering. University of Genova, 16145, Italy*

[3]*Laboratory MOSIL National School of Applied Sciences Tétouan, 93000, Morocco*

ARTICLE INFO

Article history:
Received: 22 October, 2019
Accepted: 16 December, 2019
Online: 15 January, 2020

Keywords:
Hazardous Materials
Risk
Transport
Storage
Logistics

ABSTRACT

Companies manufacturing, using or transporting hazardous materials (HM) face risks at every link in their supply chain. Industrial activities related to this type of material expose not only the workers, but also the population and environment to serious risks.
Logistic management and risk management are often considered independent. However, logistics choices related to hazardous materials mix the two aspects in an inseparable way. In fact, risk management must be synchronized with the choices and logistics strategies of companies. The analysis of the risk relating to the management of these HMs, therefore, requires making methodological choices, but also using a lot of data, which are often incomplete.
The achievement of the present study was a real challenge owing to the lack of information and statistics on the use, transportation or storage of HMs in Moroccan territory.
The purpose of this study is to provide a portrait of the industrial practices surrounding the management of hazardous materials in Morocco in order to identify the determining factors, specifically those related to health, safety at work, in the logistic choices of companies as well as the different strategies that take into account the variables: storage, transport and quantity, including cost and risk factors.
In this context, and to conduct a survey on the logistic choices related to hazardous materials at our country, we have set up a questionnaire that will try to reach the largest number of Moroccan companies. This questionnaire focuses on the practices surrounding: procurement, shipping, storage, outsourcing, risk management, the influence of economic factors on the management of hazardous materials and the elements affecting specifically workers.

1. Introduction

The number of industrial companies producing, using, storing and, or transporting Hazardous Materials (HM) is constantly increasing worldwide due to the growth of demand in various sectors [1]. Because of their nature, these materials require special attention. HM accidents are rare events. Yet their occurrence can result in catastrophic industrial consequences. Here we can take the example of chemical accidents that are defined as the release of notable amount of toxic materials during storage, production, transportation, use and disposal of chemicals. Such accidents may lead to a real disaster. For instance poisoning which affects in a serious way all the people the properties and the environment [2,4].

In general, risk is well defined in terms of both the likelihood of the incident and the magnitude of the loss, injury, and damage as a measure of economic loss, human injury, or environmental damage [5,6]. To define a risk able to manage aid decision makers in the transport of dangerous goods, it is useful to refer to a classification hierarchy of the decisional levels that may be associated with the management of that type of transport taking into account economic, environmental, and risk aspects [7,8].

*Abdellatif Soussi, Email : abdellatifsoussi90@gmail.com

The risk of transporting HM varies depending on the type of substance. Transporters of hazardous goods must be aware of how these materials are classified to ensure compliance with marking, labeling, placarding, and shipping paper requirements [9]. HMs may be classified in any of the following: explosives, gases, flammable liquids, flammable solids, oxidizing substances, poisons and infectious substances, radioactive material, corrosives, miscellaneous goods, and other regulated materials [9].

HMs transport incidents may occur at the origin or at the destination (loading and unloading) or on the route [10]. Studies have shown that the frequency of accidents during the transport of dangerous goods by road and rail has increased [11]. Most of these accidents were followed by fire, explosions and gas clouds [12,13]. As an example, statistically, there were 11,000 hazmat transport companies involving 310,000 vehicles and 1.2 million employees in China in 2015 [14].

In recent decades, several historical surveys have been published on accidents in chemical plants and in the transportation of HMs [15,18]. A survey of accidents that occurred during HM's road and rail transport shows that accidents frequency has increased in the 20th century, most of which are on the road [11]. Another survey was conducted on the practices of firms hazmat handling facilities, this research focused on hazmat shipments and mode of transportation and various HM and outsourced activities [19]. In [20] the authors conducted a survey covering both technical elements of HM such as the class of these materials, geographic sites covering activities, supply and shipping, various practices related to HM (loading / unloading) and organizational elements for companies using HMs (risk management, emergency preparedness, training, subcontractors). A survey of the accident situation of tankers carrying HMs was conducted in order to identify the different causes and hazard classes for HM accidents, consequences and corresponding probabilities [21].

The study of the literature shows that a multitude of models have developed to take into account different risks on different parts of the supply chain, to various activities [22-25]. Thus, decisions at the operational level (for example, a truck driver) will not be the same as those taken at the tactical level (for example, the head of the part of the company), nor even as those taken at the strategic level (for example, the management team). Yet, each can opt for the best decisions according to their own point of view, their own information and their own objectives. The purpose is to assure a vertical cooperation in order not to have these various objectives in comparison with one another on the decision-making plan. For instance, a decision taken at the operational level which improves the performance of the activities of the truck driver, could really have negative impacts on the global performance of the chain (channel).

Nowadays in Morocco the chemical materials are used in the most productive sectors including industry, agriculture, health, mining and quarrying and consumption [26]. However, the chemical industry is based, essentially, on three classes of products in this case: petroleum products, industrial chemicals and fertilizers.

The dangerous consequences of HM accidents have urged the Moroccan legislation to pay serious attention to the transportation of this type of materials. HMs are defined in the Moroccan law N°30-05 as any material, object or organism which, by reason of its nature, may harm the people, the properties and the environment. No research to our knowledge in this area has been conducted in Morocco to assess the overall situation.

The purpose of this study is to provide a current survey on the situation of HMs in Moroccan companies in order to identify locations, types and classes of HM, logistics chain (supply and shipping), risk management, training of the companies with regard to these materials.
Within this background, the present study aspires to (1) know the companies' methods of supplying and shipping HM (mode, frequency, and type of transport used) and the reasons behind such choices, (2) know the distribution of tasks between the shipper and the carrier during the loading and unloading operations of HM, (3) determine the different storage locations, (4) determine the HM related items representing the largest costs to companies (5) test the ability and willingness of companies to invest in risk reduction measures, (6) know the sector where subcontractors are used for operations related to HM, (7) check the companies' interest / involvement in the activities of their subcontractors, (8) find out the different risk reduction measures, (9) determine the impacts of HM accident, and finally (10) verify whether the current legislation holds industrial activities related to HM or not.

2. Materials and Methods

2.1. Survey by questionnaire

The survey's objective is to assess the companies' fixed sites and refer to logistic strategies that are adopted by them in terms of HMs. Since the two relevant factors (cost and risk) can affect such strategies, it is essential to understand the role played by these factors within these strategies. To sum up, these have been considered the hearth of the supply chain and have been assessed based on: (a) their procurement practices, (b) their expedition practices, (c) their practices on the site of their company. It is noted that the completion of this questionnaire was based on a study in Canada on the logistics choices of HMs [27].

Table 1: Companies that responded to the questionnaire and their field of activity

Companies' activity	**Numbers**
Health and pharmaceutical	3
Automotive industry	8
Chemistry and oil	8
Energy & Research	7
Aviation industry	4
Building and public works & Construction	5
Food Industry	5
Transport and logistics	3
Other industrial activities	12
Total	55

This questionnaire was sent to 103 companies that are concerned with HM from different sectors. In a period of 14 months we have received the responses of 55 companies (53.40%) as indicated in the Table I which also provides the companies' field of activity.

The questionnaire consists of various sections; each section focuses on a particular aspect of the company activities. These sections are detailed as follow:

a. Company Identification

This domain of the survey helps gather general information about the company in order to obtain a quick overview (company's name, number of employees, activity area…). And it also used to collect information related to the contact person and allows the transfer of results to the participating company.

b. Identification of a site where there is Hazardous Materials

This part of the survey asks the respondent to identify, among all the installations belonging to the company, a site where there are HM. All the following questions are specific and will be about this site. This choice was made for the sake of consistency in the responses, the nature and intensity of practices that can vary greatly from one site to another. Once the site is identified, the following information are gathered:

- The number of employees on the site;
- The geographical coverage of the site activities;
- The shared part of activities related to HMs;
- The transport of HMs classes found on the site;
- The name of three main HMs found in the site.

c. Supply of Hazardous Materials

This part of the questionnaire focuses on practices related to the supply of HMs. Supply issues have been separated from shipping issues because these practices may vary depending on these two functions of the company. Some companies supply HM but they ship very little, usually in the form of residual HMs. Questions related to supply issues focus on:

- The number of HMs received on the site;
- The frequency of the receptions;
- The mode of transport used for the supplies;
- The packaging used;
- The conditions surrounding unloading;
- The use of subcontracting.

d. Shipment of Hazardous Materials

This part of the questionnaire focuses on practices surrounding shipments of HMs. It is considered as a mirror of the previous section on supplies (c), the same questions are asked.

e. Hazardous Material at identified fixed site

This part is mainly interested in the different places of storage on the identified fixed site. More precisely, the questions can be used to determine whether the company uses temporary storage sites inside or outside its site. In addition, the questions check whether companies have HM transported frequently in order to reduce the quantities on the site.

f. HM Supply related costs

This part focuses on costs related to HMs. The questions are about:

- The criteria used by the company during these choices of supply, which makes it possible to weight the importance of the cost factor;
- The importance given by the company to various factors related to HMs (specialized vehicles, equipping, training, etc...);
- The maximum percentage increase in operating costs that the company could tolerate to invest in security measures;
- The weighting of the economic impact of different types of accidents on the company.

g. Subcontractors with activities related to Hazardous Materials

This part of the questionnaire is interested in the use of subcontracting, a common measure to most companies. The objective is in particular to verify if this practice can have repercussions on the security level. The questions asked are relate to:

- The sectors (transport, unloading, on-site handling, packaging, etc.) and where the company uses subcontractors for its Hazardous Materials activities;
- The reason which drives companies to use subcontractors;
- The level of knowledge that companies have about the activities of their subcontractors;
- The type of contracts linking companies to their subcontractors;
- The possible loss of accountability that the business suffers when using subcontractors;
- Measures to monitor the activities of subcontractors;
- The important criteria when selecting a subcontractor.

h. Risk control

This part of the questionnaire focuses on the risk management measures put in place by the companies. The questions concern:

- The risk management measures put in place on the site and during transport;
- The risk communication policy put in place by the company;
- Procedures surrounding accidents / incident management;
- The possible impacts (immediate direct cost, loss of production, image loss, etc.) of a Hazardous Material accident on the company;
- The relative importance of different types of accidents (fixed site vs. transportation).

i. Regulation Hazardous Materials

This part of the questionnaire is interested in the manner in which the companies are cooperating in accordance with the regulations on dangerous materials. Among other things, the questions check whether the different regulations (storage, transport) constrain the company's activities and whether, according to them, these regulations make it possible to manage the risk effectively.

3. Results

At this juncture; we shall present the rough results and their analysis. The first part of the chapter is dedicated to the presentation of the results obtained from the fixed sites. The second part analyzes these results and the third part introduces the differences noticed between the small and the large companies that used HMs.

3.1. Answers obtained

A total of 103 questionnaires have been supplemented to date. Four of these answers come from an intermediate version from the questionnaire which explains the high number of abstentions to the questions related to the different Residual Hazardous Materials (HMR) received or shipped on the site, frequency training offered, the maximum percentage increase in current operating costs, the economic impact on the business of an HM accident involving employees and public, personnel or departments dedicated to risk management, information or communication on the management of risks to employees, and the organization and planning of on-site activities with clients, suppliers and subcontractors.

The answers to the questions related to the Company Name and the Name of the "contact person" and to the first part of the identification of a site where there are HM are not presented, in order to preserve the identity of the guarantors. Moreover, the answers were compiled and are presented in a grouped way to obtain accurate statistics and to prevent the identifies of the companies from being exposed based on the answers.
The answers obtained come from various horizons: branches of the industry, cuts of the company and portions of the activities related to the HMs.

3.2. Analysis of the results

The following parts aim at analysing the answers obtained in the survey in order to establish bonds between the various elements and to draw essential conclusions. In this section, no distinction is made between the various classes of companies.

3.2.1. Characteristics of the responders

In this part, we shall deal with the first two sections of our survey. Namely, the identification of companies and the site where HMs are present.
The companies which have answered the questionnaire belong to different sectors, as indicated in Table 2 which also provides the companies sector of activities. Most of them belong to the group of large companies (40% have between 250 and 5000 employees), and averages companies (38.2% have between 50 and 249 employees), while a moderate percentage contains small companies (10.9% have between 10 and 49 employees) as shown in Table 3.
More than half of the guarantors (78.2%) use more than one site to carry out their activities related to HMs, these sites are divided into two types, factory or warehouse. It should be pointed out that most of the participants, (69.1%) are using the factory for these activities (30.9% using warehouse). These sites have more than one geographic coverage whether it is local, regional or national as shown in Table 4.

Table 3: Number of employees in the companies participating in the survey

Number of employees	Numbers	Valid Percentage	Cumulative percentage
Between 250 and 5000	22	40.0	40.0
Between 10 and 49	6	10.9	50.9
Between 50 and 249	21	38.2	89.1
< 10	1	1.8	90.9
> 5000	5	9.1	100.0
Total	55	100.0	

Table 4: Geographic coverage of supplies and shipments of the site

Geographic coverage	Answer	Numbers	Percentage	Valid Percentage	Cumulative percentage
Local	Yes	9	16.4	16.4	16.4
	No	46	83.6	83.6	100
National	Yes	23	41.8	41.8	41.8
	No	32	58.2	58.2	100
Regional	Yes	9	16.4	16.4	16.4
	No	46	83.6	83.6	100
Another	Yes	24	43.6	43.6	43.6
	No	31	56.4	56.4	100

When talking about Small and Medium Enterprises (SME), there is no doubt about their essential function in the Moroccan national economy. They represent the bulk of the entrepreneurial fabric, a proportion exceeding 95% and occupying more than 50% of private sector employees. The share of SMEs in Moroccan exports and domestic private investment is respectively around 31% and 51%. These entities play a vital role in promoting the social dimension as well as in the economic and territorial development.

The SME is present in all sectors of Moroccan economic activities: industry, crafts and construction, businesses and services that include tourism, communications, transport, financial services.
In the industrial sector in particular, the SME accounts for almost half of the total as follows: textiles and clothing (35%), chemistry and para-chemistry (26%), agri-food (24%), mechanics and metallurgy (12%), electrical and electronic (3%) [28].

According to the Moroccan Ministry of Industry [10], the most influential industrial sectors in the national economy are: the automotive industry for that Morocco is considered as the first construction hub on the African continent, followed by other sectors such as Aeronautics, Textile; Leather, Electronics, Electrical, Chemical-Parachemistry, Pharmaceutical, Building Materials, Renewable Energy and Mechanical and Metallurgical Industries (IMM). This explains the rations and sectors related to

Table 2: Sector of activities of the companies participating in the survey

Activity sector	Numbers	Percentage	Valid Percentage	Cumulative percentage
Health and pharmaceutical	3	5.5	5.5	5.5
Automotive industry	8	14.5	14.5	20.0
Chemistry and oil	8	14.5	14.5	34.5
Energy & Research	7	12.7	12.7	47.3
other industrial activities	12	21.8	21.8	69.1
Aviation industry	4	7.3	7.3	76.4
building and public works & Construction	5	9.1	9.1	85.5
Food Industry	5	9.1	9.1	94.5
Transport and logistics	3	5.5	5.5	100.0
Total	55	100.0	100.0	

the hazardous substances involved in this research. Not all the companies that answered the questionnaire are specialized in handling of HMs. About one fourth (21.9%) of the obtained answers comes from companies belonging to other industries whereas the other half comes from related companies of different sectors.

The Figure 1 shows the percentage of companies whose part of their activities at the site is related to HM.

We have found that there are various classes of HM on the site of the median company where most present classes are the flammable liquids (56.4%), the gases (40%), and the corrosive substances (29.1%). The rarest classes are the radioactive materials (7.3%) (figure 2). Among the various dangerous substances listed on the sites, one is generally found: gasoline, GPL (Butane), sulphuric acid and xylene.

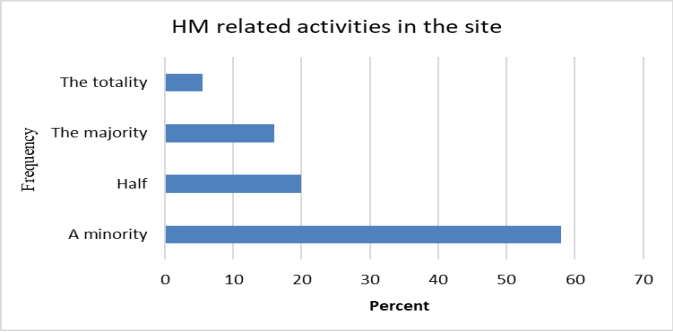


Figure 1: HM related activities in the site

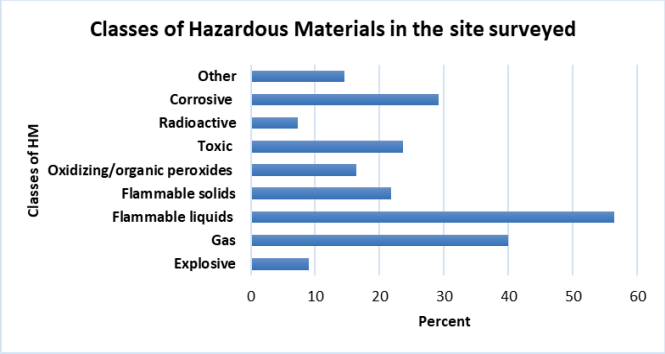


Figure 2: Classes of Hazardous Materials in the sites surveyed

Concerning the classes of HMs imported or exported through the port of Tangier Med in 2017 [29], we have noted, based on some received statistics, that the most transported classes, as shown in Figure 3, are as follows: the flammable liquids (36%), corrosive substances (19%), the gases (40%).

Concerning Class 1 (Explosive) and Class 7 (Radioactive Materials), their presence constitutes small percentage due to the special procedures used to transport the materials of these two classes, which require the approval of a special commission to be imported, exported or transported. In [30], some examples of the procedures used in the port national agency for these two classes depending on the Moroccan law have been cited.

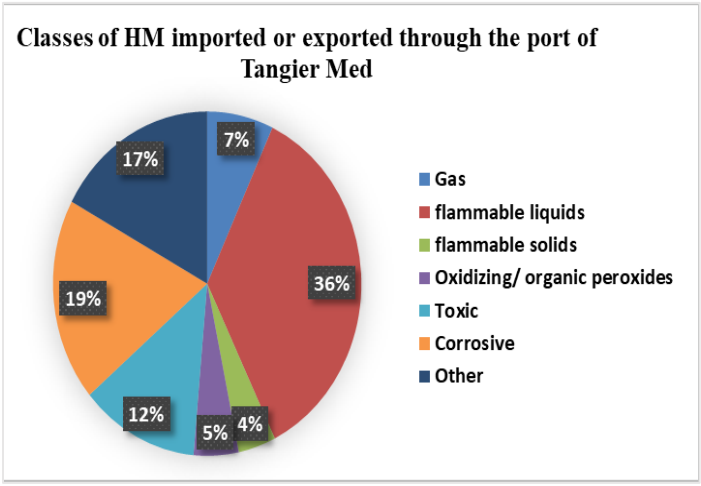


Figure 3: The classes of Hazardous Materials imported or exported through the port of Tangier Med in 2017

3.2.2. Provisioning and forwarding

The great majority of the guarantors (70.9%) affirm receiving dangerous substances or products controlled on the identified fixed site, and just 27% of the companies that make the shipment of HM or products controlled from their site as shown in the following graphic:

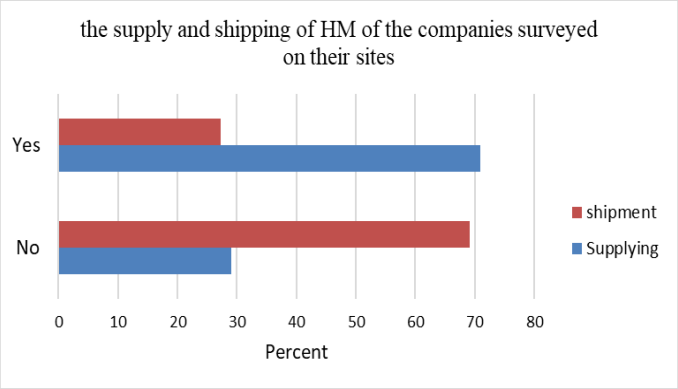


Figure 4: The supply and shipping of Hazardous Materials of the companies surveyed on their sites.

This is explained by the presence of companies consumers and non-producers of HM within the sample. The majority (27.3%) of the companies affirm the treatment of five different dangerous substances or more (figure 5). The very great majority of the companies (79.8%) receives dangerous substances during the month, the week or few times a year.

In contrast, we found the majority (10.9%) of 25.5% of companies affirming the shipment of HMs from their sites, averred that they treat five different dangerous substances or more. The very great majority of the companies (28.6%) ship HMs during the day, 23.8% during the month or a few times a month, 19% during the week (figure 6).

Figure 5: Number of Hazardous Materials received/ shipped from the sites.

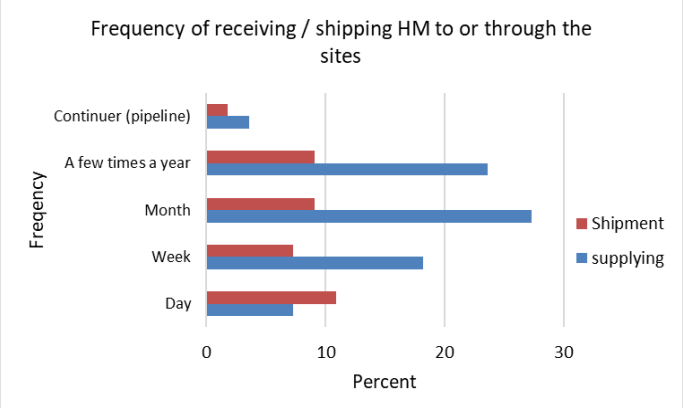


Figure 6: Frequency of receiving/shipping HM to or through the sites.

The means of transport preferred for the provisioning are, respectively (figure 7): the truck (74.5%), the train (5.5%), the boat (27.3%), pipeline (3.6%) and the plane (1.8%). The strong preponderance of the truck can surprise at first sight, but recall that according to the questionnaire, a company would notch this choice recourse to a means of transport, even if it is in an occasional way. The percentages collected cannot thus be translated into mileage, tonnage or a number of sending. We have also found that the most common mode of transportation is the conditioned mode (67.3%), which bulk transport just 18.2% of respondents who affirmed who are using its use.

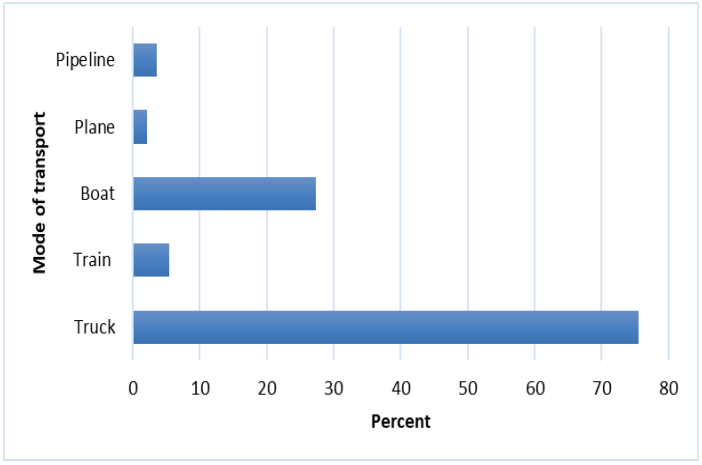


Figure 7: Mode of transportation used to transport HM

A high number of companies affirm that it would be impossible to change means of transport. The majority of the other companies affirm that they could change means of transport only for some substances but not for all the substances. It should be recalled here that the choice of a means of transport is largely influenced by the proximity of the infrastructures of transport.

Very few companies (12.7%) carry out truck transport on their own. This task is usually assigned to the supplier (41.8%) or to a third-party carrier (23.6%) (Figure 8). However, the majority of the companies supervise the unloading of the dangerous substances (41.8%) and in many cases carry out unloading (21.8%). There are on the other hand some companies that uses a subcontractor (5.5%) (figure 9).

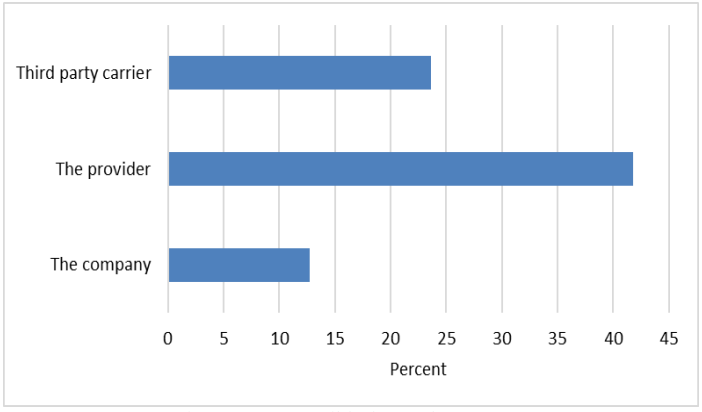


Figure 8: Responsible for truck transport.

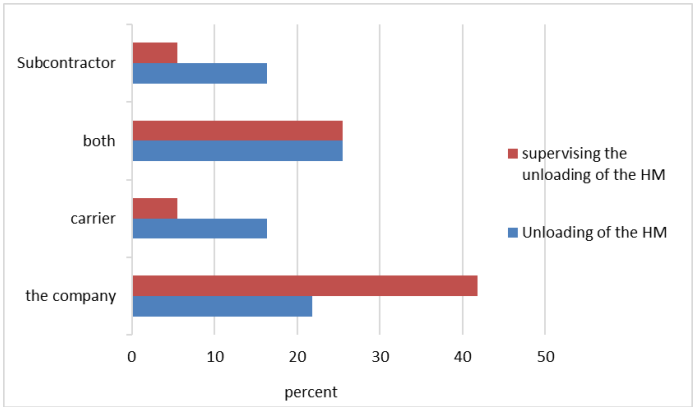


Figure 9: Supervising the unloading of HM or HMR in supply.

3.2.3. Hazardous Materials (HM) at fixed site

The training offered to the employees working with the quays of shipping/receiving is often offered by a consultant (39.2%), yet the majority (56.4%) of the employers are concerned too, as shown in Figure 10. This training is often offered during the recruitment (47.3%) or even annually (47.3%). Among the surveyed companies only (3.6%) announce that they offer this training in three years (Figure 11). The other companies include it with other trainings sessions at the time of recruitment or with the requirements.

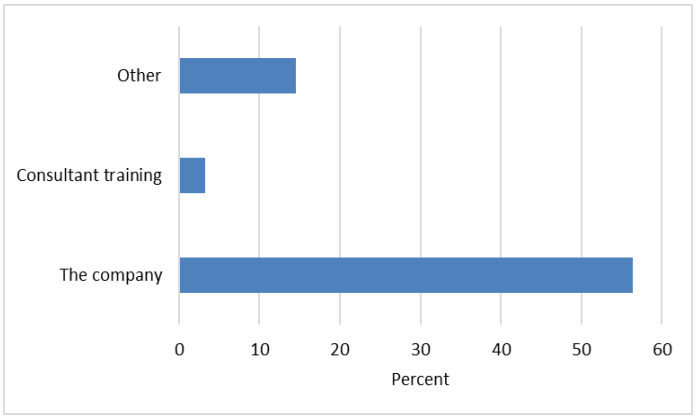


Figure 10: The training offered to the employees of the shipping/receiving.

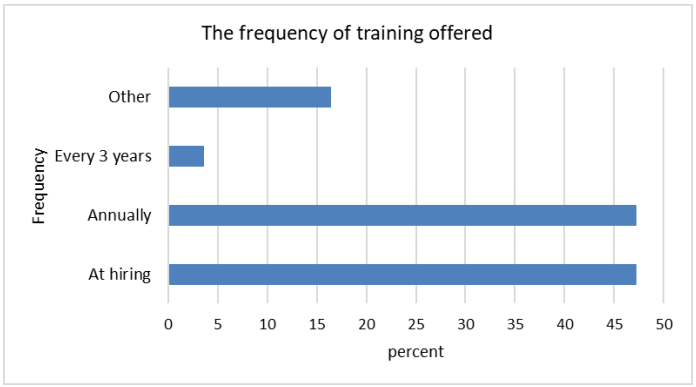


Figure 11: The frequency of training offered to the employees at site.

The majority of the companies (54.5%) always or often use zones dedicated to HM. Some companies however affirm that they're used only sometimes (10.9%) or even rarely (10.9%). In the opposite way, it is not frequent that dangerous substances are stored in a temporary way on the site of the company. A few companies (21.8%) have even stated that they often if not always have recourse to this practice.

On the other hand, temporary storage outside the site is not very widespread. Only 16.3% (3.6% always, 12.7% sometimes) of the companies use temporary storage on a site which does not belong to their own company, whereas temporary storage on other sites belonging to the company touches 18.2% of the guarantors (5.5% always, 12.7% often).

The majority of the companies affirm that they transport HM more often in order to avoid having too much of it on the site. A considerable share of the companies (65.5%) affirm that they have recourse to this practice always (12.7%), very often or sometimes (52.8%), whereas 21.8% of the companies say that they never do it, as shown in figure 12 below. This contributes to the increase in the number of convoys on the roads, although the quantities remain the same.

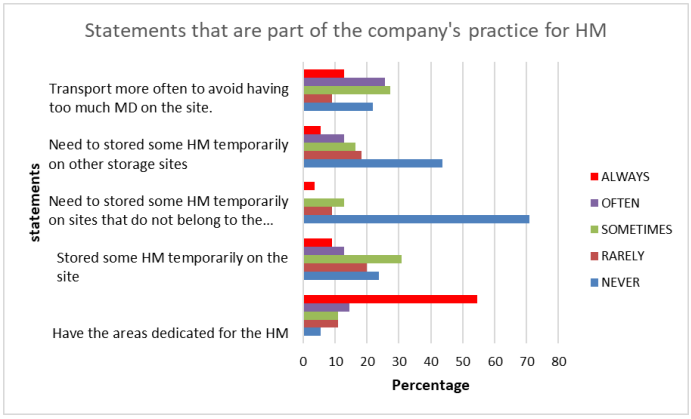


Figure 12: Statements that are part of the company's practice for HM.

3.2.4. Costs related to HM supplies and shipments

The majority of the companies (85.5%) are ready to invest more in safety since the sums do not exceed a certain level. In the vast majority of cases, this level is below 5% (for 20% of them) or even 10% (for 14.6% of companies that have answered) on current operating costs (figure 13).

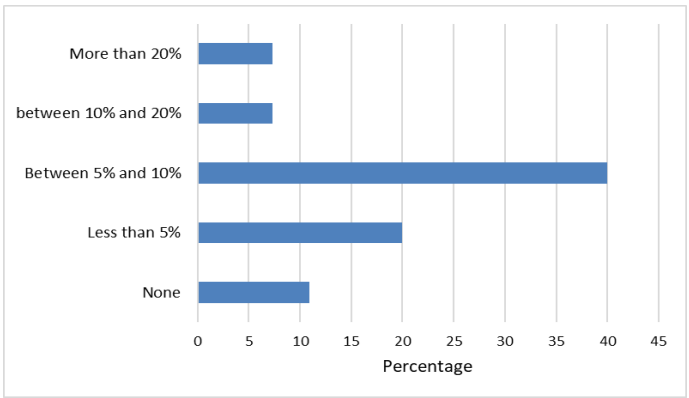


Figure 13: Percentage of current operating costs that companies would invest for security measures.

Moreover, when they make their choices of provisioning or HM forwarding, the criterion "Transport security" is mentioned very often (89.1%) followed by the "cost of transport" (83.6%) then the "cost of storage" (70.9%) as shown in figure 14.

Several companies did not classify their answers in order of importance, but even with fragmentary results, this classification offers a distribution similar to the one mentioned so far.

The most frequently mentioned HMs items when referring to costs as shown in figure 15 are, in order: equipment on the site and employee training (74.6%), security audits (69.1%), specialized vehicles (67.4%), regulatory measures (65.5%), General insurance premiums (56.4%) and hiring of a security manager (52.7%). The insurance premiums (52.7%) and the accidents/incidents HM (49.1%), are less often mentioned.

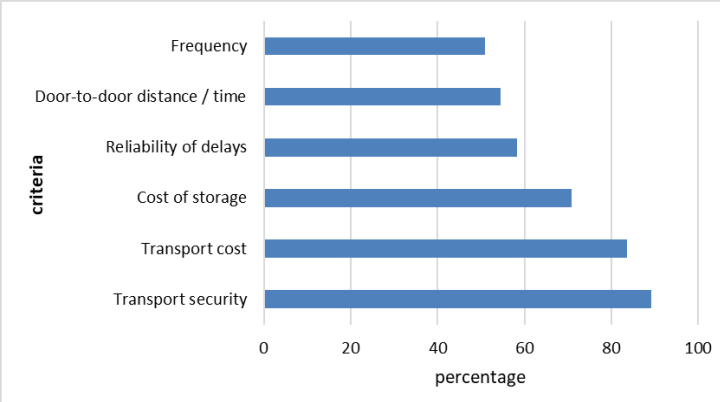


Figure 14: The criteria that are important when the company makes its choice of HM supply or shipments.

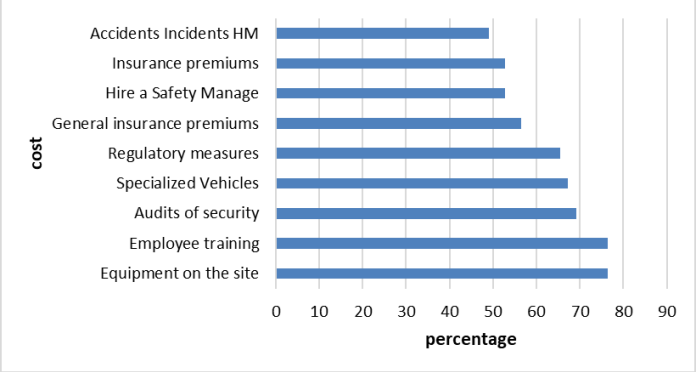


Figure 15: The costs incurred by HM represent a significant burden for the company.

The guarantors had to consider the economic impact of three types of accidents (with spill, implying employees, implying the public) compared to the economic impact of an accident isolated without discharge. In the three cases, the companies have estimated that the impact would be more important, and their evaluation of the possible impacts on their company increases with each new scenario. Indeed, 16.4% of the guarantors have estimated that the impact would be much more important for an accident with spill whereas this proportion grows to 32.7% for an accident involving employees and to 36.4% for an accident involving the public. We have noticed, however that some guarantors have estimated that the impact would be similar if the accident implied employees or the public (figure 16).

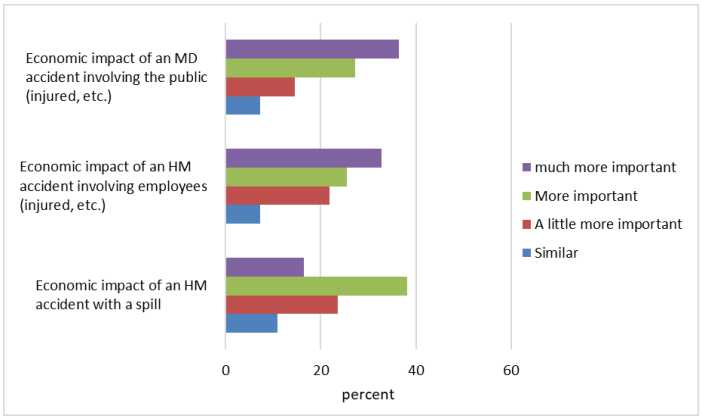


Figure 16: The economic impact of an HM accident on the company.

3.2.5. Subcontractors with activities related to HM

All companies use subcontractors for any of their HMs activities. The areas where they are most popular as shown in figure 17 are, in order: transportation (45.5% supply and 34.5% shipping), loading (40.5%) and unloading (34.5%). On the other hand, a few companies use the services of subcontractors on their site, whether while handling (27.3%), packaging (25.6%), storing (18.2%) or producing (18.2%).

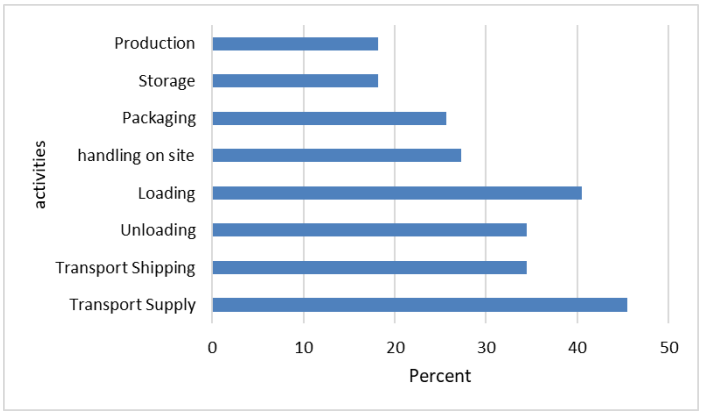


Figure 17: Subcontractors in HM activities.

Regarding the use of subcontractors for transport, companies mainly cite: the fact that they do not have the expertise (23.6%) and the fact that the carrier shares the responsibility of the risk (25.5%), the costs (16.4%) and the fact that they do not have vehicles (9.1%) as illustrated in figure 18. Conversely, the most cited reasons for not using subcontractors are risk management (32.7%) and costs (14.5%) or other reasons (23.6%).

A number of elements point to a tendency towards disempowerment. For example, several respondents have claimed that the impact of a transport accident made by their subcontractor would have a zero impact on their company (20%) or lower (25.5%) than an accident, with their own vehicles. Besides, 25.5% have said that they never perform safety audits of their subcontractors while only 20% have stated that they always do. When we combine these few elements with

the fact that 16.4% of the companies know perfectly well that their subcontractors delegate in their turn to subcontractors (29.1% ignore it and 5.5% are vaguely aware of it), one can easily realize the extent of the phenomenon of disempowerment in transport.

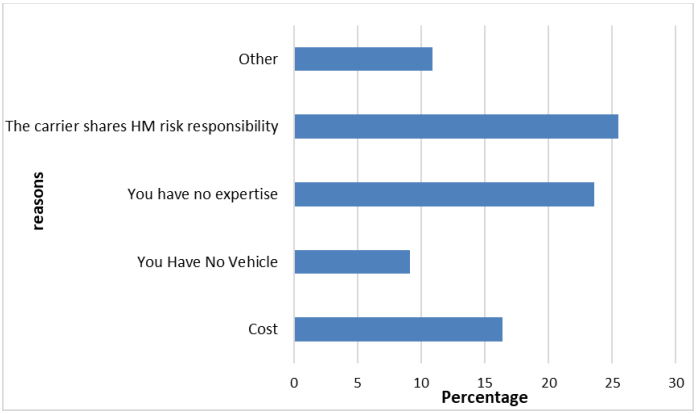


Figure 18: Reasons for using the subcontractors in the transport activities.

Regarding the contracts of these companies with their subcontractors, 10.9% of respondents still use long-term contracts with their subcontractors while 21.8% have mentioned that they never use them (Figure 19). When it comes to the selection of the subcontractors targeted by these contracts, companies are mainly concerned with: reliability / quality of service (72.7%), control of safety (70.9%), cost (69.1%), ISO certifications (60%), the ability to track transportation (52.7%), carrier reputation (50.9%) and past accidents (50.9%). The possibility of establishing a lasting relationship (45.5%) receives less attention, although they are also considered important by many companies (figure 20).

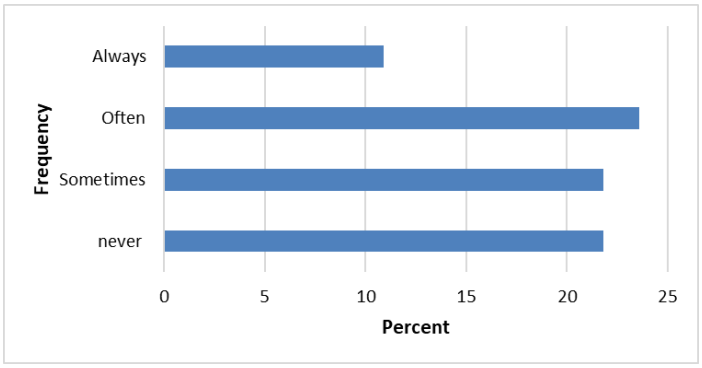


Figure 19: Long term contracts between the company and their subcontractors.

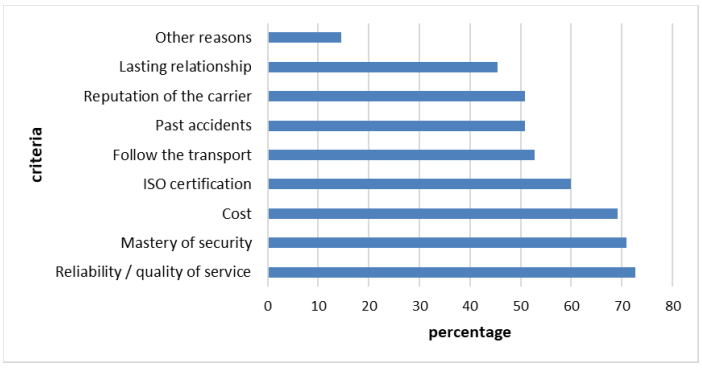


Figure 20: The relevant criteria while selecting a carrier for HM.

3.2.6. Risk management

Almost all companies report performing risk analyses on their site, in spite of numbers such as never (7.3%) rarely (7.3%) or sometimes (16.4%). On the other hand, they are less likely to do risk analyses for transportation; a number of companies never do (16.4%), rarely (10.9%) or sometimes (21.8%) and only 21.8% always do, unlike 34.5% in fixed site. This is because a large number of companies outsource transportation activities. Similarly, companies are more likely to have stricter procedures than the law on their site for transport (23.6% still use more stringent procedures than the law on the site, against 16.4% for transport). Additionally, few companies (18.2%) use tracking technologies (GPS, etc.) to transport HMs (40% never use them) (figure 21).

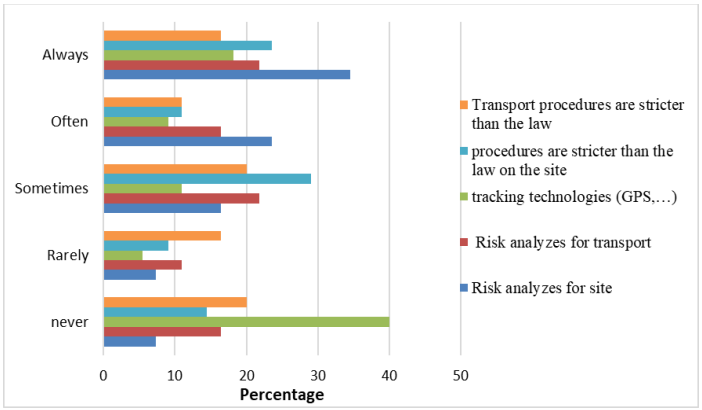


Figure 21: Risk management for the surveyed companies.

A few companies communicate their risks with citizens, whether for their site (12.7% always do it) or for transport (7.3% always do), but many companies use their risk management to enhance their image with 29.1% of respondents who have said that they never do it, as shown in figure 22.

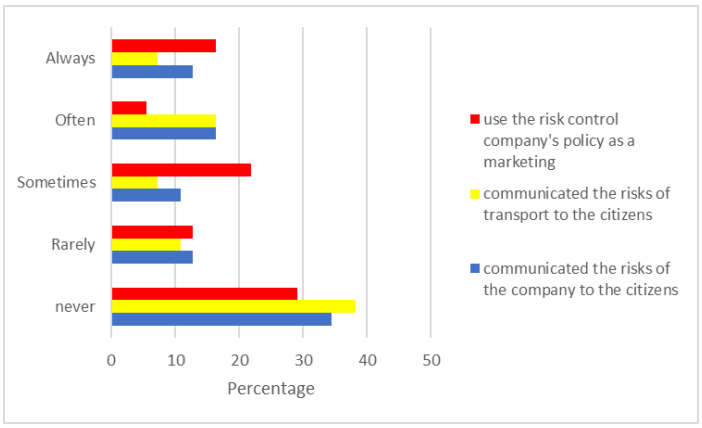


Figure 22: Risk communication in the surveyed companies.

The majority of respondents (52.7%) still claim that they use specific procedures while loading and unloading HMs. Though, some companies never use them (16.4%), rarely (3.6%) or sometimes (16.4%). On the side of the various risk management programs, we have observed the same situation. Most companies offer risk management information/communication sessions to their employees (69.1%), have a workplace health and safety

committee (69.1%), and have dedicated risk management staff (61.8%), or a prevention program specific to HM (61.8%) (figure 23).

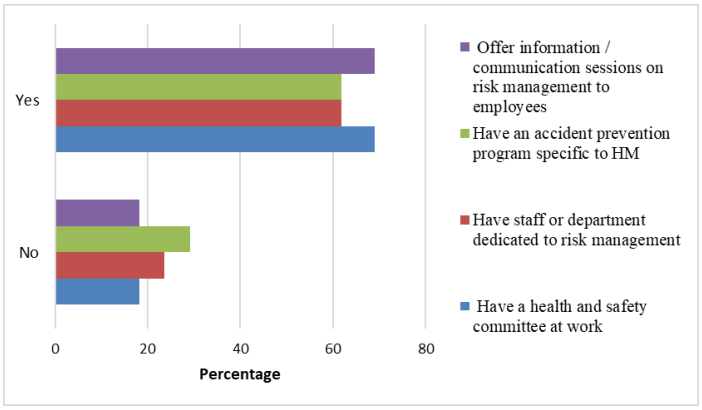


Figure 23: Risk management programs in the surveyed companies.

When the companies are questioned about the possible impacts of an HMs accident (in transport or on the site) on their company, the companies answer respectively, as shown in figure 24: the direct cost (70.9%), the loss of image (65.5%), the loss of production (65.5%), the loss of customers (60%), the increase in insurance premiums (58.2%) and the reaction of citizens (45.5%).

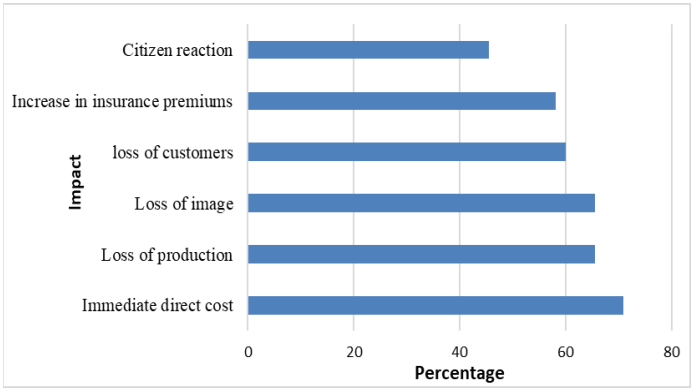


Figure 24: The possible impacts on the company in case of HMs accident.

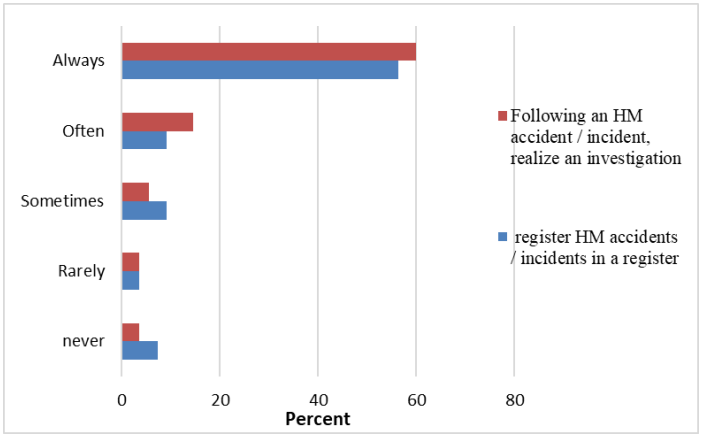


Figure 25: Accident/incident management for the surveyed companies

Most companies (56.4%) still list these accidents in a register, although 7.3% of respondents say they never do so and 3.6% rarely do. In the same way, 60% of the companies always carry out investigations following these accidents whereas 3.6% of the companies affirm that they never do it (Figure 25).

In the opinion of many companies, a fixed-site accident (49.1%) would have a greater impact than a transport accident (18.2%), as shown in figure 26.

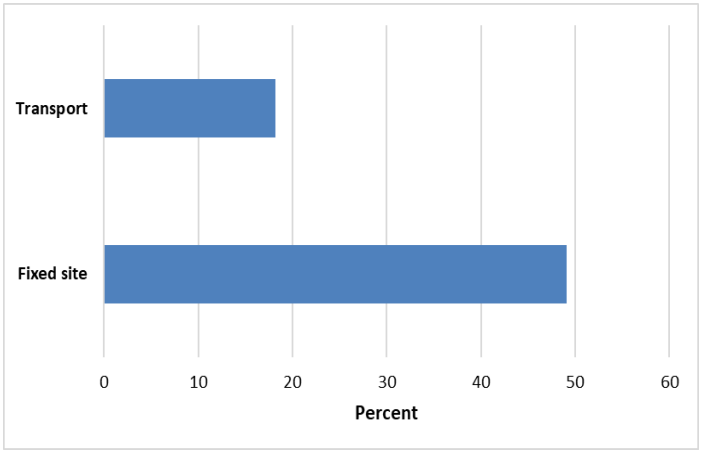


Figure 26: Type of HM accidents and their weight on the companies.

3.2.7. HM Regulation (Law No. 30-05 on the transport of dangerous goods by road)

Many companies claim that HM regulations (site, transportation, health and safety) restrict their day-to-day activities with only 23.6% of companies saying they never limit them. Storage regulations seem to be the one that restricts businesses the most. On the other hand, 30.9% of companies say that Health and Safety at Work regulations never limit them and 29.1% of companies say that transport regulations never limit them (figure 27). It should be remembered, however, that the majority of companies do not perform the transportation themselves, which may influence their perception of transportation regulations.

Due to the limitations imposed by the storage regulations, several companies had to: make substitutions of materials (54.6%), reduce their stocks of HMs (61.9%) or change their logistical choices and delivery frequencies (52.7%) (figure 28).

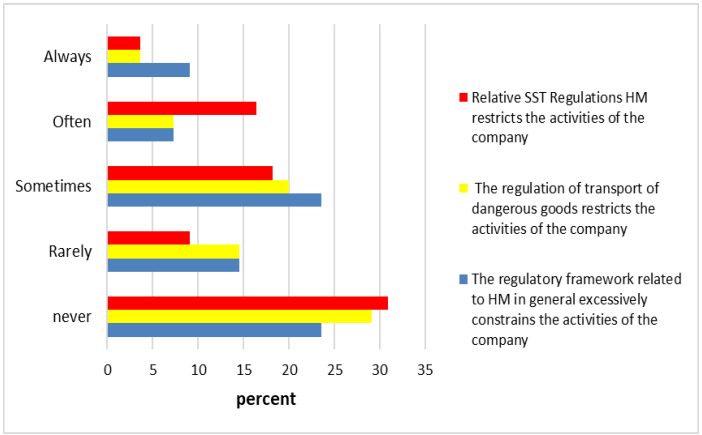


Figure 27: The affirmations related to the HM regulation according to the surveyed companies.

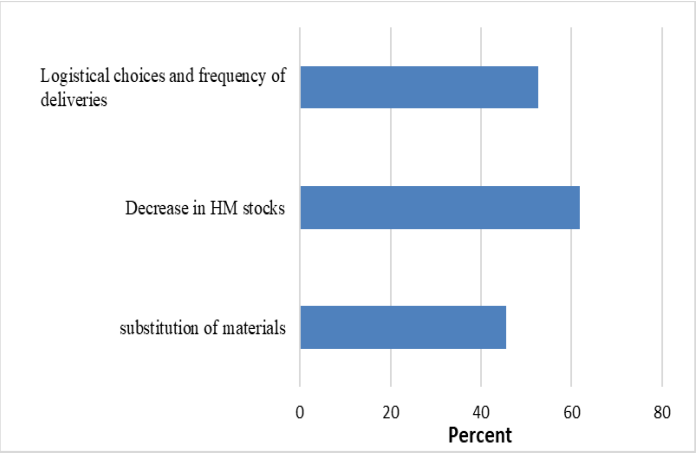


Figure 28: The choice of modifications that can be made by companies according to HM storage regulations.

4. Discussion

Our primary objective is based on the development of this part of the study by examining the nature, quantity and routes of HM in circulation in Morocco. Unfortunately, while we were doing this research, there was a lack of information and statistics on HMs even among the services and state agencies. As a result, it was difficult to formulate a coherent and precise idea on this question at the national level. That can be attributed to several reasons, the most important of which is the absence of an exhaustive official inventory of HM currently circulating in Morocco, only a few fragmentary statistics are available. They are carried out in the places of production and consumption without indicating the precise nature of the substances transported.

The classification adopted by Morocco is that developed by the UN. When examining the data from our survey, we can conclude that the classes of the most dangerous substances transported in Morocco are flammable liquids, gases and corrosive substances (figure 4), so the transport administration must be aware of the danger that major accidents that may occur, such as:

- fires and / or explosions caused by flammable materials;
- serious disorders related to inhalation or ingestion of certain toxic substances etc.

The industrial sector ranks second in the Moroccan economy [31], so the range of dangerous products, which are mainly manufactured products in circulation, is not very wide. The hydrocarbon sector and its derivatives are the ones that accounts for the majority of safety concerns related to hazardous substances transported. This sector is important on the national scale, with two respects:

- the preponderant and vital role of hydrocarbons in the country's economy;
- the largest volume of hazardous substances carried is that of hydrocarbons and their derivatives.

Hazardous substances imported or produced and distributed in Morocco mainly use the road to reach their destination (Figure 9), the mode which represents more than 75% of the national freight [32].

The analysis of accidents caused by HMs, as indicated in [17] have shown that these accidents are significantly highly occurring during the transport of theses HM, followed by accidents occurred in processing plants and storage facilities. In contrast, most domino accidents are found in storage area, followed by process plants and transportation [13], which are mainly caused by LPG, petrol and gasoline [13].

Road transport of hazardous substances has the advantage and the convenience of directly linking the producer and the consumer. In this case, the quantities transported packaged or in bulk, are relatively less important (a few tons to a few tens of tons). This also allows, at least in theory, to reduce the risk of accidents by minimizing the handling and unloading-loading operations. Moreover, users of dangerous substances can be scattered throughout the national territory. In fact, the majority of companies that produce or use as raw material the dangerous substances are concentrated in Casablanca and its surroundings and also the region of Tangier. Nevertheless, this is not a safety advantage as accidents in urban and suburban areas can be particularly disastrous.

The fact that Morocco has a fairly large network of national roads and interurban highways is an asset for the road transport of dangerous substances. Most of the accidents, in which they are involved, are resulted from vehicle failures. More precisely from the imprudence of the transport professionals.
Among the dangerous products transported are flammable liquids, gases and corrosive substances, which is represented in [21] as the most frequent hazmat tanker accidents. The rarest materials are radioactive materials (Figure 2).
Even if the increase in the number of tankers used to transport oil and other dangerous substances is justified by the ever-increasing demand for these products, which are essential for economic activity, the fact remains that this activity generates a multitude of potential dangers (fires, explosions, intoxications, negative impacts on the environment).

The constraints of competitiveness and productivity imposed on industrial, mining and agricultural stakeholders, mean that the transportation of these products to the exploitation sites is carried out at a frantic pace and generally without much precaution. Moreover, the multiplicity of the risks of pollution and the various other dangers that this field of transport provokes is often unpredictable, to the point that in the event of an accident, these risks took the appearance of a catastrophe not only for the environment but also for users (the populations).
For example, the accidental spillage of flammable, explosive or highly toxic products is always the most dangerous because the damage caused would be disastrous.

Unlike the road network which covers the entire national territory, the rail network for the transport of HM is limited to Atlantic Morocco with two axes serving the Northeast (Oujda) and Central West (Casablanca) respectively. A good part of this network is reserved for the transport of phosphates.

With the recent development of passenger transport, the Moroccan railway does not offer a large place for goods in general, and even less for dangerous goods. The ones it is

accustomed to and continues to transport are always connected to phosphate (phosphoric acid, sulfuric acid, fertilizer). Explosives are also among the dangerous substances carried by the railways. The administration of ONCF applies strict regulations on them, and no major accident has ever occurred.
The transport of dangerous goods by rail was regulated by the Dahir n° 1-01-223 of 30 August 2001, confirming the ratification by Morocco of the Convention concerning International Carriage by Rail (COTIF) and its appendix on the transport of dangerous goods better known. also, under the initials RID [33].

In terms of risk management, the results show that 47.3% of companies received HMs training for the entire supply chain during recruitment and thereafter on a yearly basis (Figure 10). So, the majority of companies expressed their willingness to supply security equipment and almost all companies have shown that the choice of HM supply or shipments is based primarily on the criteria of transport security (89.1%) and transport cost (83.6%) as shown in Figure 14.

The data analysis also highlighted some specific characteristics. Most accidents are due to operations such as the loading / unloading of HM as indicated in [13], these operations as shown in our research (Figure 17) are dedicated mainly to the subcontractor, it seems that an additional effort should be devoted to improve the security of these procedures, including stricter legislation and better training.

It is reminded that the decree of application of the law 30-05 relating to the road transport of dangerous goods was published in the Official Bulletin on June 30, 2011 [30]. The law fixes the conditions of security imposed on the carriers and envisages sanctions and sentences ranging from one month to two years and fines of DH 2,000 to DH 100,000. In order to align with European standards for road transport of dangerous goods, more precisely with European Agreement on International Transport of dangerous goods by road, known as ADR. Law 30-05 determines all the conditions applicable to this type of transport: from loading to unloading, from the conditions of use of the vehicles to the obligations incumbent on the operators in the operation.

Though Morocco is among the first signatory countries to these agreements, the application of these agreements is not yet widespread because of lack of implementing legislation. As a result, with the exception of oil companies and multinationals, most tanker drivers do not distinguish between a potato load and radioactive or explosive material. Several Moroccan companies organize ADR training for the benefit of their drivers and can travel in Europe without any problem to transport dangerous goods on European soil. There is no point in passing a law if it is not applied in the field. What we are seeing now is that some customers are not demanding and always favor the cost factor at the expense of security.
Several accidents caused by tanker trucks had serious economic, social and environmental consequences. Since 2007, the Bouregreg Watershed Agency has sued two carriers for accidentally spilling part of their oil load. The first directly in the wadi Bencheqcheq and the second, near the wadi Amiran in the region of Bouznika. "*The cleaning operations cost 220,000 dirhams and we ask the authors of the pollution damages*" [34].
Nowadays, there is a real gap between multinationals whose number remains limited and local businesses which are likely to be more numerous. Multinationals with significant resources are more compliant with safety standards for the transport of dangerous products and substances. Local operators, on the other hand, are relatively less sensitive to the importance of these standards. But, we still lack application texts. In the absence of these texts, professionals do not apply fortiori the regulatory provisions. By contrast, multinationals apply the standards of their parent companies while local companies do not apply the regulations. The change in this situation that threatens to give rise to a disaster at any time in the event of an accident requires a real awareness on the part of professionals and decision-makers.

5. Conclusion

This study presents the results of a survey on logistic choices related to HMs in Morocco. it focused on the determination of the nature of these HMs, in terms of classes of materials transported, frequency of supply and shipment, etc, and then identified the most used organizational security practices by Moroccan companies.

This study undertakes to determine the situation of hazardous materials in Morocco, we tried to give an analytical and holistic analysis for the whole transport logistic chain of these HMs.

Organizational and technical security measures need to improve through the introduction of a culture of safety throughout the HM chain of shippers, handlers, carriers and consignees of HMs, who must therefore receive training adapted to their field of activity and their level of responsibility especially for classes (classes 3, 2 and 8) as well as the high presence in the Moroccan industry sector and which have the highest risk in case of accident.

The results obtained in this research can help the interests involved (authorities, professionals and companies) for strategic and operational planning of the transport chain of dangerous products in Morocco.

The exploratory research we have conducted can pave the way for a wide variety of development perspectives. It might be interesting to extend this survey to the kingdom as a whole in order to highlight possible discrepancies in the use of organizational security practices.

Among other things, the possibility of developing a guide to good practices for carriers to inform them of the practices they can put in place to reduce their risk, or the possibility of making many recommendations to different ministries in relation to the training, the use of equipment, accident register policies, or even mount communication campaigns to raise awareness of the risks of multi-client carriers and multiple loading / unloading operations.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors thank all the companies and people who participated in this work for their great support and valuable cooperation.

References

[1] J. Yang, F. Li, J. Zhou, L. Zhang, L. Huang, and J. Bi, "A survey on hazardous materials accidents during road transport in China from 2000 to 2008," *J. Hazard. Mater.*, vol. 184, no. 1–3, pp. 647–653, 2010.

[2] B. Y. Kara, "Designing a Road Network for Hazardous Materials Transportation," *Transp. s*, vol. 38, no. 2, pp. 188–196, 2004.

[3] J. Zhang, J. Hodgson, and E. Erkut, "Using GIS to assess the risks of hazardous materials transport in networks," *Eur. J. Oper. Res.*, vol. 121, pp. 316–329, 2000.

[4] V. Verter and E. Erkut, "Hazardous materials logistics : An annotated bibliography," *Oper. Res. Environ. Manag.*, pp. 221–267, 1996.

[5] Center for Chemical Process Safety (CCPS), *Guidelines for chemical process quantitative risk analysis*. Center for Chemical Process Safety of the American Institute of Chemical, 2000.

[6] A. Soussi *et al.*, "Risk analysis for hazardous material transport by road : case study on Tangier-Tetouan region , Morocco," *018 13th Annu. Conf. Syst. Syst. Eng. (SoSE). IEEE, 2018.*, pp. 464–470, 2018.

[7] C. Bersani and C. Roncoli, "Real-time risk definition in the transport of dangerous goods by road," *2012 7th Int. Conf. Syst. Syst. Eng.*, 2012.

[8] C. Bersani, C., Minciardi, R., Tomasoni, A.M., Sacile, "Risk averse routing of hazardous materials with scheduled delays.," in *Security and Environmental Sustainability of Multimodal Transport, NATO Science for Peace and Security Series C: Environmen*, 2010, pp. 22–37.

[9] G. Rechkoska, R. Rechkoski, and M. Georgiska, "Transport of dangerous substances in the Republic of MAcdonia," *Procedia - Soc. Behav. Sci.*, vol. 44, pp. 289–300, 2012.

[10] E. Erkut, "Hazardous Materials Transportation," in *Handbook in OR & MS*, vol. 14, no. 06, 2007, pp. 550–552.

[11] A. Oggero, R. M. Darbra, M. Mu, E. Planas, and J. Casal, "A survey of accidents occurring during the transport of hazardous substances by road and rail," *J. Hazard. Mater.*, vol. 133, pp. 1–7, 2006.

[12] A. Conca, C. Ridella, and E. Sapori, "A Risk Assessment for Road Transportation of Dangerous Goods_ A Routing Solution," *Transp. Res. Procedia*, vol. 14, pp. 2890–2899, 2016.

[13] R. M. Darbra, A. Palacios, and J. Casal, "Domino effect in chemical accidents : Main features and accident sequences," *J. Hazard. Mater.*, vol. 183, no. 1–3, pp. 565–573, 2010.

[14] X. Huang, X. Wang, J. Pei, and M. Xu, "Risk assessment of the areas along the highway due to hazardous material transportation accidents," *Nat. Hazards*, vol. 93, no. 3, pp. 1181–1202, 2018.

[15] B. Fabiano and F. Currò, "From a survey on accidents in the downstream oil industry to the development of a detailed near-miss reporting system," *Process Saf. Environ. Prot.*, vol. 90, no. 5, pp. 357–367, 2012.

[16] J. Planas-Cuchi, E., Montiel, H., & Casal, "A survey of the origin, type and consequences of fire accidents in process plants and in the transportation of hazardous materials.," *Process Saf. Environ. Prot.*, vol. 75, no. 1, pp. 3–8, 1997.

[17] J. Vilchez, J. A., Sevilla, S., Montiel, H., & Casal, "Historical analysis of accidents in chemical plants and in the transportation of hazardous materials," *J. Loss Prev. Process Ind.*, vol. 8, no. 2, pp. 87–96, 1995.

[18] G. F. List, P. B. Mirchandani, M. A. Turnquist, and K. G. Zografos, "Modeling and Analysis for Hazardous Materials Transportation : Risk Analysis , Routing / Scheduling and Facility Location," *Transp. Sci.*, vol. 25, no. 2, pp. 100–114, 1991.

[19] M. Leroux, N. De Marcellis-warin, and M. Trépanier, "Safety management in hazardous materials logistics," *Transp. Lett.*, pp. 13–25, 2010.

[20] M. Peignier, I., Leroux, M. H., de Marcellis-Warin, N., & Trépanier, "Organizational safety practices of hazardous materials carriers Organizational safety practices of hazardous materials carriers," *Transp. Lett.*, vol. 3, no. 3, pp. 149–159, 2011.

[21] X. Shen, Y. Yan, X. Li, C. Xie, and Lihua Wang, "Analysis on Tank Truck Accidents Involved in Road Hazardous Materials Transportation in China Analysis on Tank Truck Accidents Involved in Road Hazardous Materials Transportation in China," *Traffic Inj. Prev.*, vol. 15, pp. 762–768, 2014.

[22] A. Ditta, O. Figueroa, G. Galindo, and R. yie-pinedo, "A review on research in transportation of hazardous materials," *Socioecon. Plann. Sci.*, 2018.

[23] K. Van Raemdonck, C. Macharis, and O. Mairesse, "Risk analysis system for the transport of hazardous materials," *J. Safety Res.*, vol. 45, pp. 55–63, 2013.

[24] M. Verma, "Railroad transportation of dangerous goods : A conditional exposure approach to minimize transport risk," *Transp. Res. Part C*, vol. 19, no. 5, pp. 790–802, 2011.

[25] E. Erkut and A. Ingolfsson, "Transport risk models for hazardous materials : revisited," *Oper. Res. Lett.*, vol. 33, no. 1, pp. 81–89, 2005.

[26] IMIST (Institut Marocain de l'Information Scientifique et Technique), "Bulletin d'Information Technologique Chimie / Parachimie," Maroc, 2012.

[27] M. T. et I. P. Nathalie de Marcellis-Warin, *Stratégies logistiques et matières dangereuses*. 2013.

[28] B. EL MAITIA, "memoireonline," *l'Organisation et le Financement des Marchés Publics au Maroc*, 2008. [Online]. Available: https://www.memoireonline.com/05/08/1118/m_organisation-et-financement-marches-publics-maroc28.html?fbclid=IwAR2bh7atVh62Oz8fCSd8RiZeXMPMwcTcJCHYZO-uqjiR8BHHSnBlZF_Ik1w.

[29] TMSA, "Tangier Med Special Agency," 2017. [Online]. Available: www.tmsa.ma/en.

[30] Gouvernment of Morocco, "Dahir n° 1-11-37 du 29 joumada II 1432 (2 juin 2011) portant promulgation de la loi n° 30-05 relative au transport par route de marchandises dangereuses." Bulletin officiel, Morocco, 2011.

[31] HCP (Haut Commisariat du Plan, Morocco), "Situation économique nationale 2018," 2018.

[32] Ministry of Economy, Morocco, "Le secteur de transport des marchandises : Contraintes et voies de réformes," 2013.

[33] Moroccan government, "Dahir n° 1-01-223 du 10 joumada II 1422 (30 août 2001) portant publication de la Convention relative aux transports internationaux ferroviaires (COTIF), faite à Berne le 9 mai 1980." Bulletin officiel, pp. 468–541, 2002.

[34] Hassan EL ARIF, "Réforme du transport L'anarchie continue pour les matières dangereuses," 2014. [Online]. Available: https://www.leconomiste.com/article/917230-r-forme-du-transportl-anarchie-continue-pour-les-mati-res-dangereuses. [Accessed: 06-Nov-2018].

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 1, 34-39 (2020)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

Advanced Hybrid Energy Harvesting Systems for Unmanned Aerial Vehicles (UAVs)

Cuong Van Nguyen[1], Toan Van Quyen[2], Anh My Le[2], Linh Hoang Truong[2], Minh Tuan Nguyen [*2]

[1]*Thai Nguyen University of Information and Communication Technology, Vietnam*

[2]*Thai Nguyen University of Technology, Vietnam*

ARTICLE INFO

Article history:
Received: 19 November, 2019
Accepted: 30 December, 2019
Online: 20 January, 2020

Keywords:
Unmanned Aerial Vehicle (UAVs) - Drones
Solar cells
RF energy harvesting
Hybrid energy harvesting
Rechargeable batteries

ABSTRACT

The recent use of rotary-wing unmanned aerial vehicles (UAVs) has gained significant interest and continuously been implemented since they are used across the world for civilian, commercial, as well as military applications. The drawback of UAVs is the fight-time due to the limited battery capacity. Therefore, energy harvesting (EH), captured energy from the ambient environment is one of the effective measures to prolong the flight time for UAVs. This paper proposes the hybrid EH system, which can simultaneously harvest power from solar and radio frequency (RF) energy sources to significantly improve the energy issues for endurance longer flight UAVs. A 7-stage voltage multiplier circuit of the stand-alone RF-EH system is designed and simulated in this work. The stand-alone solar harvester is a solar panel. The DC output voltage from both of two energy sources is passed through a DC-DC boost converter and stabilizer. Simulation results can be deployed to power for the battery of UAVs in practice.

1 Introduction

Recently, the use of flying platforms such as unmanned aerial vehicles (UAVs), popularly known as drones, is rapidly growing for many different applications in all commercial, civilian and military fields. The UAVs are very common with real-time monitoring, providing wireless coverage, remote sensing, search and rescue, delivery of goods, security and surveillance, precision agriculture, etc. UAVs have several key potential applications in wireless systems with their inherent attributes such as mobility, flexibility, and adaptive altitude. They can be equipped with cameras, sensors, and others to be more likely to perform various tasks such as surveillance and military services [1], search, and rescue in disaster damage [2], remote sensing [3, 4]. Additionally, they play an important role in the telecommunications infrastructure supporting the Internet of Things (IoT) vision [5] and wireless sensor networks (WSNs). Commercial UAVs commonly use the lithium and lithium-ion batteries that can maintain a fight time of about 20 to 40 minutes [6]. However, the energy limitations are a problem not only for UAVs but also for wireless sensor networks [7]. When UAVs do not use any alternative energy sources from the ambient environment, they may fail to several tasks due to the limited operating time. To complete various tasks for UAVs, the sustainable sources of power are considered as an effective solution. Energy harvesting techniques are the process by which energy is derived from ambient sources such as solar energy, wind energy, vibration, electromagnetic energy, etc. They can be employed to extend the lifetime of battery for low power devices. However, this paper has taken into account the performance of hybrid RF- Solar EH system which is generally a workable script, ensuring for operation UAVs, it continues during all-day. Solar energy is harvested by solar panels and then converted into electrical energy. RF energy harvesting from ambient environments is responsible as an alternative power source for powering the day and night flight when there is no irradiation and thus no power can be extracted from solar panels. Both solar and RF energy sources contribute to the system with the same role. Hence, we propose a hybrid system which comprises of the RF energy harvesting and on-board solar cell for the UAVs long-endurance flight. Rechargeable batteries get support from both sources to operate the UAVs. To make sure the UAVs can avoid suddenly cutting power, the output power of the hybrid system is utilized for UAVs through the battery. All the circuits for either stand-alone or combined methods are designed with analysis and simulation results to clarify our methods. The hybrid method shows promising points to be more likely to improve the energy issues for UAVs in practical. In summary, the contributions of our paper are:

[*]Corresponding Author: Minh Tuan Nguyen; Tel: +84-208-3847-093; Email: nguyentuanminh@tnut.edu.vn; tuanminh.nguyen@okstate.edu

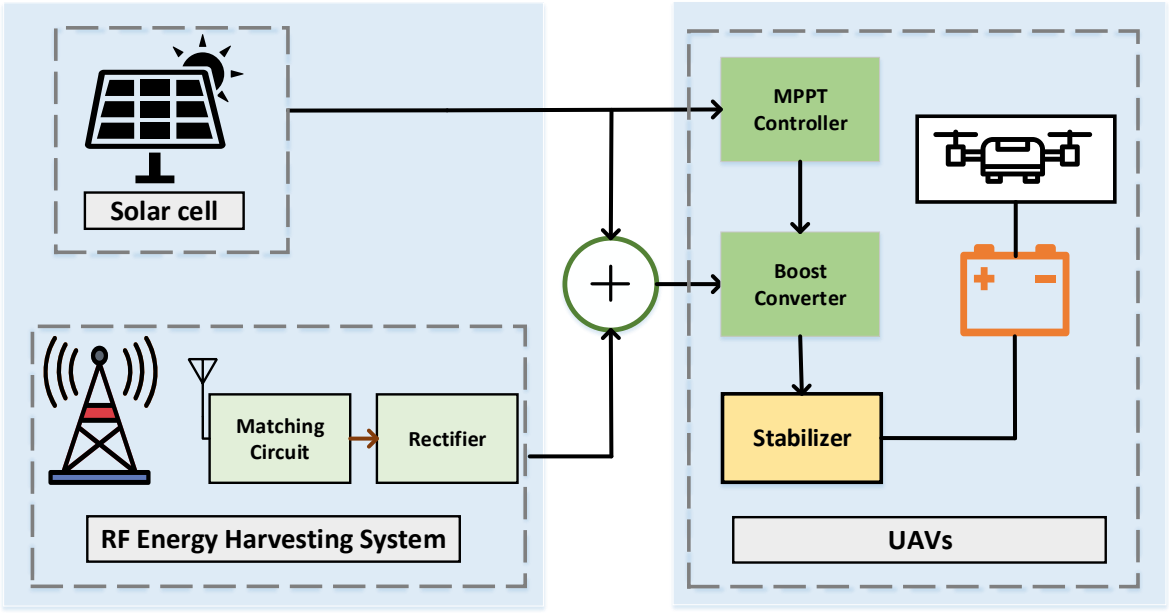


Figure 1: The block diagram of hybrid RF-Solar system for UAVs

1. A new model of energy harvesting for UAVs is proposed.
2. Analysis and Design with real circuits are provided for clarification.
3. Recommendations in choosing suitable design including circuits, electronic elements to be able to achieve the most effective are provided.

The remainder of this paper is addressed as follows. All the related work is provided in Section 2. Theoretical Analysis and Design is mentioned in Section 3. Simulation results are provided with separated methods, RF energy, and solar energy harvesting, and combined both of them in Section 4. Conclusions and future work finally addressed in Section 5.

2 Related work

Radio-frequency (RF) energy can be collected from the environment, called ambient RF energy harvesting (RF-EH) which considered as free powering sources. Therefore, the utilization of the perpetual ambient RF-EH is more attractive than other dedicated energy sources of the WPT system. RF-EH system consists of three main components: (1) the receiving antenna, (2) the impedance matching circuit and (3) the voltage multiplier (VM). However, due to the low power density of the RF signal, the DC output voltage of the VM could be not as high as enough to supply energy for charging mobile electric devices (MEDs) such as UAVs. In order to overcome the above drawbacks, several methods are proposed for improving the VM circuit topologies in which the aim is improving conversion efficiency from ambient RF power to direct-current(DC) voltage at the given frequency band to power the low power devices/circuits such as the design of multiband and broadband rectifier [8], the choice of the number of stages [9, 10], etc.

Besides, the RF/DC efficiency and DC output voltage is the proportion with the number of stages [11]. For examples, an equivalent incident signal of 0 dBm, a 5 stage voltage doubler circuit in [12] achieved V_{out} = 1.8 V while Devi *et al.* in [13] used 7 stages Schottky diode voltage doubler circuit with the result obtained is DC output voltage of 5.0 V. In the other hand, in wide-bandwidth RF applications that require smaller form factors or reduced cost, direct RF energy source is a suitable choice. Similar to the ambient RF system, a direct RF system also needs to improve the efficient RF-DC conversion system. The existed solutions applied for charging battery UAVs can be mentioned as designing a dual-band rectifier with impedance matching [14], designing a light-weight flexible rectenna array mounted on the UAV body micro-controller drove power management circuit that enables efficient battery charging [15], etc. Another energy source, the solar energy considered as harvesting energy of the WPT system, is the cleanest and most plentiful renewable energy source available. Solar panels are used to absorb solar energy and convert into electric power through the photovoltaic (PV) effect. Therefore, solar energy harvesting technologies are widely used for various applications, especially generating electricity for domestic, agriculture, and industry, etc. In reality, the solar system poses many issues in terms of environmental conditions, initial high capital cost, low efficiency, etc. To eliminate these above issues, the efficiency of the solar cell system can be improved by applying methods such as a maximum power point tracking (MPPT) techniques [16, 17], a solar power management system (SPMS) [18] and boost converter in [19, 20] which can get the high efficiency with slight weight. Besides, we also need to choose the most suitaestble materials for manufacturing solar panels. In [21], thin-film solar cells is fabricated with Cu_2ZnSnS_4 (CZTS) absorber layers. The battery is the storage place for UAV energy. Wireless power transfer is the most popular technology to charge for the battery of UAVs. This technology satisfies high-performance power trans-

mission over short to long distances with various methods such as mganetic resonant coupling, laser beaming, etc. [22].

Harvester aforementioned circuits can be developed to collect energy from stand-alone either RF or solar source. In this paper, we proposed a hybrid RF-solar harvesting system simultaneously harvests energy from solar and RF sources readily available in the ambient environment for the purpose of charging UAV batteries.

3 Theoretical Analysis and Design

This section provides an analysis of each stand-alone source system and also the design to combine suitable components to build the hybrid system. All the designs will be simulated to clarify the methods and also the recommendations.

3.1 RF energy harvesting

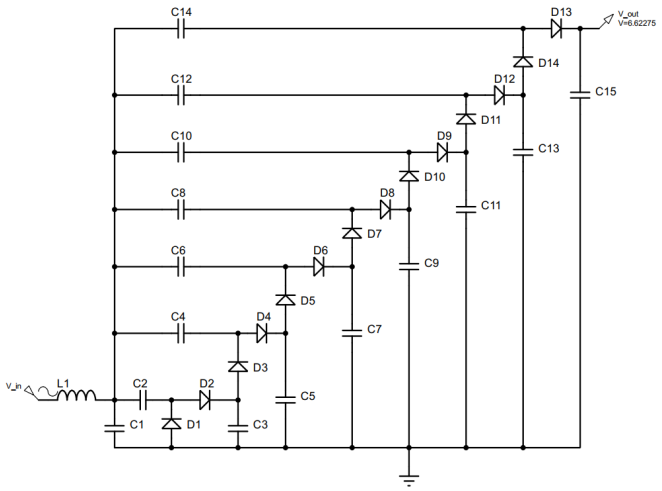


Figure 2: RF energy harvesting

This model includes three main components, antennas, matching circuit and rectifier. After receiving RF signal by antenna, a matching circuit is installed in series between RF signal source and the rectifier to be able to mitigate the dissipated power for these parts. The rectifier circuit converts the input of an AC voltage into a DC output. In each rectifier stage, the voltage multiplier method is used to increase the harvested output voltage from narrow frequency input [10]. The seven-stage voltage multiplier (VM) circuit is designed and simulated to be shown in Figure 1. The 7-stage VM circuit consists of passive components such as stage diodes (D_1-D_{14}) and stage capacitances (C_1-C_{14}). Each stage is stacked onto the previous stage.

3.2 Solar enery harvesting

Solar panels are deployed to convert light from the sun into electricity using semiconducting materials that exhibit the PV effect. They have disadvantage sides of varied efficiency value under weather conditions. In Proteus software [23], if its sunny bright day then solar panel normally gives in the range of 15 V to 19 V. On a cloudy day, it could vary between 8 V and 12 V. Solar panels obtain output ranges from 2 V to 6 V in the night time. In general, choosing solar panels could affect the output voltages. In other words, depending on the desired voltage needed for each UAV system, we can choose a suitable one for the system. This also depends on the intensity of sunlight in a deploying area.

In this paper, in order to keep the UAV system light and active, the number of cells should be limited. In addition, we want to improve the efficiency of solar battery charging in the night time that motivates us to choose the output voltage as 5 V. Based on this chosen voltage level, the solar panels will be chosen accordingly. Therefore, DC-DC boost power converters are also chosen in the systems to step up the output voltage of a solar panel for a given set of conditions and maximize the efficiency of stand-alone solar energy. The circuits will be provided in the next section.

3.3 Boost Converter and Stabilizer

This boost converter is a DC to DC converter with an output voltage greater than the input voltage. In this converter, an inductor is used to resist any changes in current by creating and destroying a magnetic field. While a transistor switch is typically a MOSFET to pulse width modulate the voltage into an inductor. Maximum power point tracking controls the duty cycle of the MOSFET's switching in order to achieve the desired power value. Figure 3 illustrates a circuit diagram used for the Proteus simulation of a boost converter.

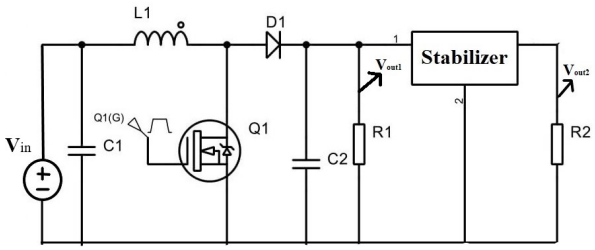


Figure 3: DC-DC Boost converter and Stabilizer

The key principle of the boost converter can be divided into two states in one cycle: state 1 and state 2. State 1: when MOSFET Q_1 is switched on at time t_{on}, the input current rises and flows through inductor L_1 and MOSFET Q_1. State 2: when MOSFET Q_1 is switched off at t_{off} = (1 - t_{on}), the input current flows through L_1, diode D_1, C_2, and load R_1. The energy stored in inductor L_1 flows through the load [24].

$$\frac{(V_{in} - V_{Trans})t_{on}}{L_1} = \frac{(V_{out1} - V_{in} + V_D)t_{off}}{L_1} \quad (1)$$

At t_{off} = 1 - t_{on}, the output voltage through the resistor R_1:

$$V_{out1} = \frac{V_{in} - V_{Trans}t_{on}}{1 - t_{on}} - V_D \quad (2)$$

Here V_D is the voltage drop across the diode D_1, and V_{Trans} is the voltage drop across the transistor MOSFET Q_1. V_{out1} is the output voltage across the resistor R_1.

Since the output voltage through the load R_1 is unstable, the stabilizer is added after the load to fix the output voltage of the

load. Therefore, the purpose of this circuit is to measure the output voltage V_{out2} across the resistor R_2.

3.4 The Design of Hybrid Energy Harvesting System

This paper proposes a hybrid RF-Solar system designed to harvest energy from two energy sources: Solar and RF energy. In order to maximize UAVs flight endurance, the conventional power supply of UAVs is integral with RF and solar energy harvesting systems. As shown in Figure 1, our hybrid system comprised of two main components, that is the energy sources which include RF and solar energy and amplifier circuits which include MPPT controller, boost converter and stabilizer. Solar cells absorb sunlight as a source of energy to generate DC electricity. Following in the solar cell design of the solliance thin film solar research [25] in which they combine a semi-transparent Perovskite cell and a CIGS cell to achieve the power conversion efficiency of 23% on a flexible solar cell.

RF energy is collected by an antenna through the matching circuit to ensure the maximum power delivery to voltage multiplier (VM). The RF power is converted from a lower AC/RF voltage to a higher DC voltage by voltage multiplier. A voltage multiplier (VM) is a special case of the rectifier circuit, including capacitors and diodes. By comparing the results of experiments [10, 12, 13], and the trade-off between the high voltage and power loss, the 7-stage voltage multiplier is chosen.

Due to the changing temperature and variation of irradiation, the efficiency of solar panels is low. A maximum power point tracking (MPPT) circuit is used for two purposes. The first is to find the amount of sunlight index harvested by solar cells and MPPT circuits extract the maximum achievable power from the PV generator. The second is support boost converter circuit to adjust the panel DC output voltage to the desired value and the duty cycle of the DC-DC converter. Two DC power sources are fed into the boost converter which is a basic power converter that operates by simultaneously closing and opening of a high frequency switches like MOSFET to step-up voltage. However, it has a drawback such as the oscillating voltage. The stabilizer is used to fix the output voltage value which supplies the battery of UAVs. Our paper assumes all parameters of the drone M200 V2 and its intelligent fight battery (TB55) that is composed of 6S Li-Po cells that correspond to a battery voltage of 22.8 V, with a capacity between 7660 mAh [26].

4 Simulation Results

In this section, we independently simulate RF and solar system and then combine them as the hybrid system to show the improvement. The UAV/Drone, named M200-V2 is chosen for our simulations. Based on that, all the circuits are designed to support the UAV. All the simulation results would be provided to be suitable for the attached batteries.

4.1 Solar Energy Standing Alone

As mentioned above, we assume the output voltage of the solar panel in Proteus simulation is 5 V which is delivered directly to the boost converter. The output voltage of the solar system with the boost converter significantly increases compared to the original value. After 40 sec, the DC voltage increases from 5 V to 15.3189 V and the current from $5e^{-12}$ A to 0.6963 A, as shown in Figure 4. Since the output voltages from the boost are unstable, a stabilizer is used to provide a stable and secure power supply to the devices.

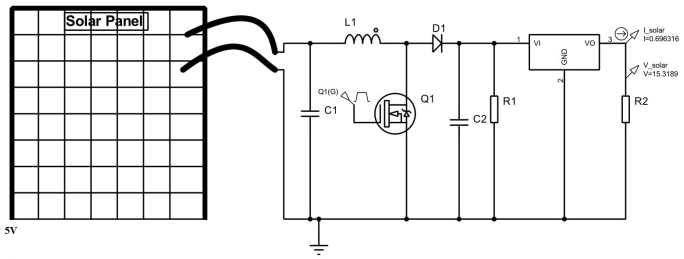


Figure 4: The solar panel connected with DC-DC boost converter

This allows the system to keep the voltage stable and to protect the systems from almost the main problems that may happen. With all of the effort to improve the solar systems, we only achieve the output voltages that are smaller than the voltages required by the drones as 22,8 V. Therefore, the hybrid energy harvesting system should be proposed to be able to adapt the power supply for the batteries of UAVs.

4.2 RF Energy Harvesting Standing Alone

In this work, the DC output voltage obtained through simulation as 6.62V, shown in Figure 2. This result is comparatively much better than in reference [13].

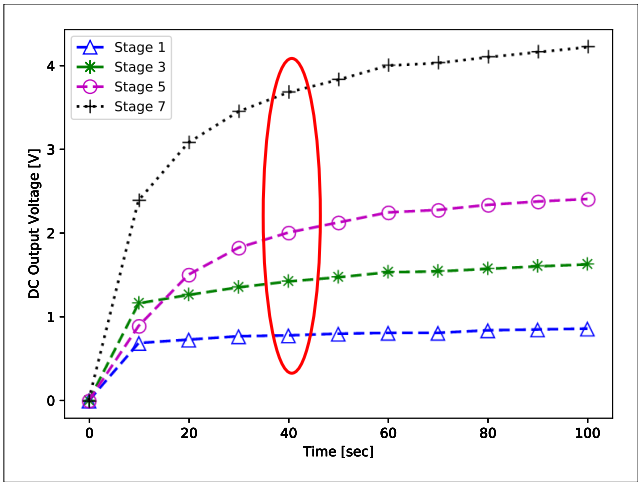


Figure 5: DC output voltage verses rise time of voltage multiplier circuit through the stages

The number of stages in the rectifier has a dramatic influence on the DC output voltage of the RF energy harvesting circuit, shown in Figure 3. Settling time more than 100 second, the DC output voltage is obtained V_{output} = 0.85 V at one stage;V_{output} = 1.62 V at 3-stage; V_{output} = 2.4 V at 5-stage; and V_{output} = 4.22 V at 7-stage. Because of the trade-off between the value of the DC output voltage and the time response, we choose the stopping time at 40 sec and

achieve V_{output} = 3.68 V with 7-stage. At this point, the voltage output is quite stable which would be enough to support the UAVs.

4.3 The Proposed Hybrid Energy Harvesting System

This simulation is aimed to take the suitable output voltage for the battery requirements of the UAV M200-V2.

As shown in Figure 6 and Figure 7, the charging voltage and current of the battery reach a stable level after one second. The one second can be acceptable because this is the response time to charge the battery rather than supply power directly to the UAV, so it does not require strictly the real-time. After that, the voltage and current values are stable at 23.15 V and 1.05 A. These results also overcome the previous systems that show promise.

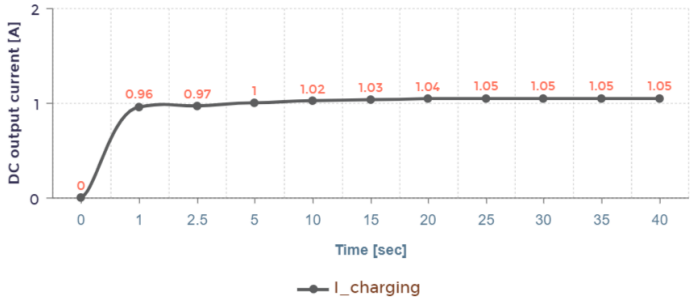


Figure 6: Time response of DC output current for the battery of UAVs

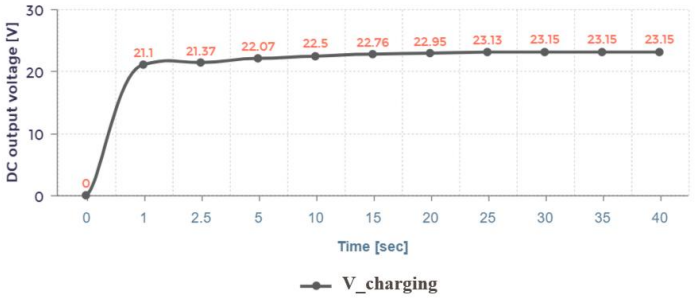


Figure 7: Time response of DC output voltage for the battery of UAVs

From the analysis of these simulations, the measured DC voltage and currently collected from RF energy, solar energy, and hybrid models are addressed in Table 1. The output current and voltage of the hybrid system are always greater than the output current and voltage of stand-alone RF or solar harvester circuit before adding the boost converter.

Table 1: The output current and voltage of Solar, RF and Hybrid system without a boost converter

	Solar Energy	RF Energy	Hybrid
Current (A)	$5e^{-12}$	$3.677e^{-8}$	0.32
Voltage (V)	5	3.68	7.17

As shown in Table 1, both the voltage and the current are not big enough to supply to the battery of the UAV/Drone M200-V2. Therefore, we have Table 2 where the circuit adds the boost converter and a stabilizer. The results show the desired levels that we always expected. We finally achieve the level suitable for the systems.

Table 2: The output current and voltage of Solar, RF and Hybrid system with a boost converter

	Solar Energy	RF Energy	Hybrid
Current (A)	0.69	0.48	1.05
Voltage (V)	15.3	9.77	23.15

5 Conclusions and Future Work

In this paper, the proposed hybrid solar-RF energy harvesting overtakes stand-alone solar and RF harvesting systems and shows promise. At any time during the day, the efficiency of the hybrid harvest system is critically increased more than stand-alone systems, so the flight time of UAVs is significantly extended. Since the stand-alone RF and solar systems are abruptly drop, so this performance of UAVs significantly improve the situations. By utilizing the hybrid system, this problem is tackled. A new design with all results is shown to clarify the problems.

Currently, the capacity of the battery storage has not been mentioned that would be the next step to improve the systems. Besides, it is possible to improve the efficiency and reduce the overshooting time by using fuzzy or PSO (particle swarm optimization) algorithm for controlling the duty cycle of the DC-DC boost converter. Further, we would utilize this hybrid source to directly supply for the UAVs and to charge for the battery of the UAVs at the same time.

6 Acknowledgement

This work is supported by Thai Nguyen University of Technology (TNUT), Vietnam, project T2018-B40.

References

[1] P. Lewis, "Cctv in the sky: police plan to use military-style spy drones," *The Guardian*, vol. 23, p. 1, 2010.

[2] S. Chandrasekharan, K. Gomez, A. Al-Hourani, S. Kandeepan, T. Rasheed, L. Goratti, L. Reynaud, D. Grace, I. Bucaille, T. Wirth, *et al.*, "Designing and implementing future aerial communication networks," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 26–34, 2016.

[3] U. Niethammer, M. James, S. Rothmund, J. Travelletti, and M. Joswig, "Uav-based remote sensing of the super-sauze landslide: Evaluation and results," *Engineering Geology*, vol. 128, pp. 2–11, 2012.

[4] E. R. Hunt, W. D. Hively, S. Fujikawa, D. Linden, C. S. Daughtry, and G. McCarty, "Acquisition of nir-green-blue digital photographs from unmanned aircraft for crop monitoring," *Remote Sensing*, vol. 2, no. 1, pp. 290–305, 2010.

[5] N. H. Motlagh, T. Taleb, and O. Arouk, "Low-altitude unmanned aerial vehicles-based internet of things services: Comprehensive survey and future perspectives," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 899–922, 2016.

[6] B. Lee, S. Kwon, P. Park, and K. Kim, "Active power management system for an unmanned aerial vehicle powered by solar cells, a fuel cell, and batteries," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 4, pp. 3167–3177, 2014.

[7] M. Nguyen, H. Nguyen, A. Masaracchia, and C. Nguyen, "Stochastic-based power consumption analysis for data transmission in wireless sensor networks," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 6, no. 19, 2019.

[8] A. Collado, S.-N. Daskalakis, K. Niotaki, R. Martinez, F. Bolos, and A. Georgiadis, "Rectifier design challenges for rf wireless power transfer and energy harvesting systems," *Radioengineering*, vol. 26, no. 2, pp. 411–417, 2017.

[9] G. DiCataldo and G. Palumbo, "Design of an nth order dickson voltage multiplier," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 43, no. 5, p. 414, 1996.

[10] P. Nintanavongsa, U. Muncuk, D. R. Lewis, and K. R. Chowdhury, "Design optimization and implementation for rf energy harvesting circuits," *IEEE Journal on emerging and selected topics in circuits and systems*, vol. 2, no. 1, pp. 24–33, 2012.

[11] R. E. Barnett, J. Liu, and S. Lazar, "A rf to dc voltage conversion model for multi-stage rectifiers in uhf rfid transponders," *IEEE Journal of solid-state circuits*, vol. 44, no. 2, pp. 354–370, 2009.

[12] T. R. Ansari, A. Khan, and I. Ansari, "Wireless charging of mobile battery via optimization of rf energy harvesting system," *International Journal of Scientific & Engineering Research*, vol. 6, no. 7, 2015.

[13] K. K. A. Devi, N. M. Din, and C. K. Chakrabarty, "Optimization of the voltage doubler stages in an rf-dc convertor module for energy harvesting," *Circuits and Systems*, vol. 3, no. 03, p. 216, 2012.

[14] K. Suri, M. Mohta, and A. Rajawat, "Design of a dual band rectifier circuit for drone powering applications," in *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–4, 2017.

[15] S. Dunbar, F. Wenzl, C. Hack, R. Hafeza, H. Esfeer, F. Defay, S. Prothin, D. Bajon, and Z. Popovic, "Wireless far-field charging of a micro-uav," in *2015 IEEE Wireless Power Transfer Conference (WPTC)*, pp. 1–4, IEEE, 2015.

[16] M. Kerdegari, "Optimization and design in control of intelligent solar uav using mppt," in *Iranian Aerospace Society Conference*, 2012.

[17] E. Sener, I. Turk, I. Yazar, and T. H. Karakoç, "Solar powered uav model on matlab/simulink using incremental conductance mppt technique," *Aircraft Engineering and Aerospace Technology*, 2019.

[18] J.-K. Shiau, D.-M. Ma, P.-Y. Yang, G.-F. Wang, and J. H. Gong, "Design of a solar power management system for an experimental uav," *IEEE transactions on aerospace and electronic systems*, vol. 45, no. 4, pp. 1350–1360, 2009.

[19] R. Patel, S. K. Sinha, *et al.*, "Performance analysis of pv system integrated with boost converter for low power applications," in *Advances in Interdisciplinary Engineering*, pp. 879–890, Springer, 2019.

[20] A. Azeem, M. K. Ansari, M. Tariq, A. Sarwar, and I. Ashraf, "Design and modeling of solar photovoltaic system using seven-level packed u-cell (puc) multilevel inverter and zeta converter for off-grid application in india," *Electrica*, vol. 19, no. 2, pp. 101–102, 2019.

[21] J. J. Scragg, P. J. Dale, and L. M. Peter, "Towards sustainable materials for solar energy conversion: Preparation and photoelectrochemical characterization of cu2znsns4," *Electrochemistry Communications*, vol. 10, no. 4, pp. 639–642, 2008.

[22] M. T. Nguyen and T. H. Nguyen, "Wireless power transfer: A survey of techniques, and applications on communication networks," 2018.

[23] "Solar panel library for proteus online: theengineeringprojects.com/2018/07/solar-panel-library-for-proteus.html."

[24] P. W. Chan, "Dc-dc boost converter with constant output voltage for grid connected photovoltaic application system," in *Industrial Electronic Seminar*, 2010.

[25] "23% efficiency for flexible cells – for real?, https://www.solliance.eu/2019/23-efficiency-for-flexible-cells-for-real/."

[26] T. Campi, S. Cruciani, and M. Feliziani, "Wireless power transfer technology applied to an autonomous electric uav with a small secondary coil," *Energies*, vol. 11, no. 2, p. 352, 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 40-47 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A Blockchain-Based Architecture for Smart Healthcare System: A Case Study of Saudi Arabia

Saleh Albahli[1*], Rehan Ullah Khan[1], Ali Mustafa Qamar[2]

[1] *Department of Information Technology, College of Computer, Qassim University, Saudi Arabia*

[2] *Department of Computer Science, College of Computer, Qassim University, Saudi Arabia*

ARTICLE INFO

Article history:
Received: 09 October, 2019
Accepted: 04 January, 2020
Online: 15 January, 2020

Keywords:
Healthcare
Blockchain
Smart healthcare
Smart cities
Healthcare cloud

ABSTRACT

eHealth is the use of Information and Communication Technologies (ICT) to enhance health quality access. Health care innovations are an essential element of Vision 2030 and National Transformation Program (NTP) 2020 operational plan that will lead to an improvement in the quality of health care in Saudi Arabia. The objective of this paper is to address the healthcare system limitations in the Kingdom of Saudi Arabia in general and the Qassim region in particular to global standards. In practice, the creation of an infrastructure for storing and seamless sharing of health data between different entities is studied. Furthermore, an in-depth analysis of current practices w.r.t. the health data, as well as finding similarities between NTP 2020 and Health 2020 (European Union) has been performed. A multi-level blockchain eHealth system is proposed to provide seamless electronic health records to the patients. Moreover, the current practices being employed in Qassim province, KSA has been analyzed. It was found that private hospitals give access to medical reports to their patients besides allowing them to manage their appointments. On the other hand, access to the government hospitals' medical records is minimal.

1. Introduction

The thrust of technology has provided a lot of facilities for everyday life. The technology transformation in the medical domain evolves within the political, administrative and socio-economic factors of a particular country. Also, the technology introduced by using various platforms including e-Health, Telemedicine, and m-Health provides the facility to disseminate health-related information distributed among clinics, health care providers and patients. e-Health is a subset of telemedicine and m-Health [1-3]. The importance of these state-of-the-art innovative approaches towards improved health care systems and quality of life in the Kingdom of Saudi Arabia (KSA) is highlighted through a number of project calls initiated by the health ministry. Also, the KSA government stresses that the country's healthcare institutions must work as an allocative and collaborative body for the well-being of citizens.

In the Middle East, the KSA healthcare sector is the largest. According to the 2018 budget, the KSA spent 36% of its budget on the health sector which holds the largest share. Population growth, emerging diseases, and fluctuating international scenarios are pushing the up-gradation of the KSA health system. The government is also taking steps towards marketization and privatization as a core policy of health system reforms. With the up-gradation of traditional health to e-Health, the privacy and security of patients' data is also a significant concern [4].

The current Saudi Arabian government has initiated a Vision 2030 [5]. The Vision 2030 is a bold but achievable blueprint for an ambitious nation. For the capacity building and for achieving Vision 2030 objectives, the KSA government has launched the National Transformation Plan (NTP). The NTP vision is to create an infrastructure and clear a path for achieving the objectives of the Vision 2030.

NTP consists of 8 themes, and among the various themes, one of the most important ones is the "health care services." NTP not only focuses on quality health services, but also prioritizes the digital records growth to 70% by 2020. The government is thus focusing on implementing the electronic database of the patient's medical history timely and effectively. NTP also prioritizes the timely quality of health access to all the citizens and creates

* Saleh Albahli, Email: salbahli@qu.edu.sa

awareness about patient rights. It also provides a blueprint for effective health services to the masses taking advantage of the electronic services. It is thus an interesting direction to study the progress made by NTP in the health sector. Furthermore, we are also interested in NTP tackling the health issues compared to the European nations. These parameters thus motivated us to analyze a number of variables related to the KSA health in general and Al Qassim province in particular.

Therefore, in this paper, the healthcare system of KSA in general and Al Qassim, in particular, is analyzed with respect to global European standards. In practice, the creation of an infrastructure for storing and seamless sharing of health data between different entities is studied. This helped in performing an in-depth analysis of current practices with respect to the health data, as well as finding the similarities between NTP 2020 and Health 2020 (European Union). Finally, a multi-level blockchain eHealth system is proposed to provide seamless electronic health records to patients. Lastly, the current practices being employed in Al Qassim province are also analyzed. In the study, it was found that private hospitals give online access to medical reports of their patients besides allowing them to manage their appointments. On the other hand, access to the government hospitals' medical records is minimal, especially in Al Qassim province.

The rest of the paper is organized as follows: Section 2 presents the related work. Section 3 explains NTP 2020 and Euro Health. Section 4 presents the Smart Health Framework. Section 5 discusses different perspectives. Section 6 presents use cases and analysis of current practices, and Section 7 concludes the paper.

2. Related Work and Open Challenges

The term "Smart City" has gained considerable popularity in recent years. The concept of the smart city refers to the urban infrastructure utilizing the latest Information and Communication Technologies (ICT) for citizen welfare. In general, a smart city should have three properties: Firstly, it allows the collection of data related to various urban activities and provides a flexible system to analyze and process it. Secondly, it can find and support immediate solutions using sensors and processing units interconnected through different media. Thirdly, it combines the information from multiple modalities and thus helps in decision making, thereby, adopting an information route between the municipality and different urban communities [6, 7].

In [8] the authors have analyzed the architectural design of IBM's Intelligent Operating System (IOS) as a smart solution for cities to combine different departments.

As noted by [9], the 19th century could be regarded as an era of empires whereas the 20th century was a century of nation-states. However, the current century would be an era of cities. The cities have become a focal point of innovation. However, most of the cities are being managed under different domains with separate entities.

A number of smart cities are planned in Saudi Arabia as well. They include King Abdullah Economic City in Rabigh with an investment of 100 Billion Saudi Arabian Riyal (SAR); Knowledge Economic City in Medina with a cost of 25 Billion SAR focusing on Education and Media supplemented by real estate and tourism; Prince Abdul Aziz Bin Mousaed Economic City in Hail with an investment of 30 Billion SAR focusing on agri-business, building materials and logistics; and Jazan Economic City with an investment of 100 Billion SAR with a focus on heavy industry projects.

The Kingdom of Saudi Arabia is the largest ICT market in the Middle East by both capital volume and spending [10]. According to [11], a smart city involves the use of data along with processes that respond to the data. Most of the efforts go in top-down fashion without involving the general public. Moreover, it fails to concentrate on the citizens' needs and views them as passive consumers of city services and mere generators of data. On the other hand, citizens should be involved in the co-creation process of different services. Similarly, they have discussed a smart city project for Milton Keynes, the United Kingdom whereby smart innovative solutions are being developed keeping in view its expected growth until 2026. One of the key components of this city is its "data hub" comprising of data from energy and water consumption along with the transport.

In [12], the authors presented an integrated framework to better understand the smart cities. They identified eight key factors for the smart city concept: management & organization, technology, governance, policy context, people & communities, economy, built infrastructure along with the natural environment.

Furthermore, Health Level Seven (HL7) is an American National Standards Institute accredited Standard Developing Organization (SDO) working in the healthcare industry [13]. It deals with diverse issues revolving patients such as patient administration, patient care, results operating, etc. The standards are used for information exchange within a healthcare organization as well as between various healthcare organizations. Moreover, HL7 contains model requirements for Personal and Electronic Health Records (PHR / EHR).

In [14] the authors have discussed how smart city ICT could enable effective healthcare provision along with lowering the associated costs. They discussed how the data collected from the mobile devices could be used for personal healthcare as well as healthcare of an entire smart city [14]. Users normally search for different options for nutrition, fitness, doctors, etc. Smart cities could also employ embedded technologies such as ambient sensors to fulfill the health needs of its citizens.

Increasing pressures and demands on health systems call for a change in the way of organizing and managing the delivery of health services: The quality of service expected from the hospitals and the increasing number of patients in KSA hospitals calls for a productive change in the procedures of managing and delivering

health services. These include but are not limited to the following procedures:

- Digitization, integration of care services and including new services
- Introduction of patient-centered health care rather than doctor-centered health care
- Treating patients as customers to the hospital
- Managing shortages and distribution of health resources
- Remote accessibility of patients' records

The usage of electronic applications and tailored solutions are infesting our daily lives and helping to tackle many challenges of society. Therefore, it is believed that integrated eHealth is a part of the solution if the current technologies are used for improving health services.

Health 2020 [15] is a framework adopted by the European Union to merge the governmental approach with the societal one to improve the overall health conditions of populations. It also helps in removing the inequalities in health services, envision public health and promote people-centered health systems that are implementable, sustainable and deliver high-quality services to the people. As such, Health 2020 guides the European infrastructures and governments to have an eHealth vision, sets a strong strategic direction and provides suggestions for improving health services. It also addresses the inequalities found in health services and lays a future path.

Mobile and communication devices have penetrated the societies and many masses are using these devices excessively than ever before [15]. As an example, every person in Europe uses the Internet at least once or twice daily [14]. An interesting pattern of search has been found and which motivates the development of eHealth services. In [16] the authors showed that half the population of Europe with Internet access is using it for searching health-related tips and information. In [17], it is found that eHealth has enabled the patients to talk to doctors and medical specialists who are being available online. Therefore, it is believed that a similar trend lays ahead and is becoming a necessity in the KSA in general. As such, effective and efficient eHealth services are the need of the day. In this context, the Vision 2030 and the National Transformation Program (NTP) 2020 [5] motivates the investment in health services, focus on eHealth by increasing spending in this area. It also motivates the increase of private-sector spending from 25 percent to 35 percent. The government wishes to increase the facilitation for Saudi nurses and supporting staff to 150 per 100000 from 70.2.

To meet the policy implementation requirements, the KSA needs to push IT infrastructure and integrating the eHealth paradigm in its primary health care package, improving health care governance, improving patient safety and quality of service, and integrating data analytics in eHealth.

The current KSA vision supports the digital healthcare paradigm introduction for masses and innovations as a key for realizing the KSA Vision 2030. The eHealth integration and IT solutions will help in elevating the quality of healthcare and providing a better quality of service [18].

Based on the flexibility it provides, many health systems have relied upon the cloud to store electronic health records. This has multiple benefits: It enables seamless sharing of health records among small clinics to hospitals but also takes advantage of the reliable storage facilities of IT giants like Amazon and Google. This thus reduces the burden of the hospitals in arranging and maintaining the data centers locally thereby reducing costs and focusing on primary service of health care only [19]. Figure 1 shows the generic model for integrating eHealth for the reforms process in a society [20]. The following are the challenges for the eHealth paradigm [21].

- Lack of technical expertise and computer skills.
- Failure of adoption Health Information Services
- Human and cultural barriers
- The meaningful sharing and combining of health records data between heterogeneous systems.
- Medication safety
- Financial barriers
- Security and Privacy

3. Healthcare Initiatives: NTP 2020 and Euro 2020

Euro Health 2020 is the framework of European policy that supports government and society in the interest of health and well-being [15]. Fifty-three countries in the European Region endorsed a new evidence-based concept of health policy and value for the region which was termed as the “Health 2020”. Health 2020 focuses on improving health for all and reducing health inequality through improved leadership and management for health. The focus is on today's serious health problems. Euro Health 2020 identifies a number of priority areas focusing on asset development and sustainability in communities, empowerment, and enabling healthy environments. Euro 2020 also presents detailed guidelines for promoting public health services and improving the current health systems. Health 2020 was approved in two forms: a European policy framework that supports actions in government and society for the health and well-being of politicians and policymakers, and a health policy framework and strategy 2020 that provides more detailed details. The application of Health 2020 in related countries is currently a top priority for the Region.

Vision 2030 consists of 96 objectives, with the number of Key Performance Indicators (KPIs). Different KSA governmental bodies cooperate to achieve these objectives. A number of initiatives, jointly developed with organizational bodies with the support of the strong will of the current ruling setup are transforming many sectors in KSA. The vision also includes private, and non-profit organizations to provide input support for the new vision. The Council of Economic and Development affairs constituted an integrative governance model to achieve the objectives, and thus realize the vision [5]. For the capacity building and for achieving Vision 2030, the KSA setup has launched NTP, which involves the 24 government agencies. The

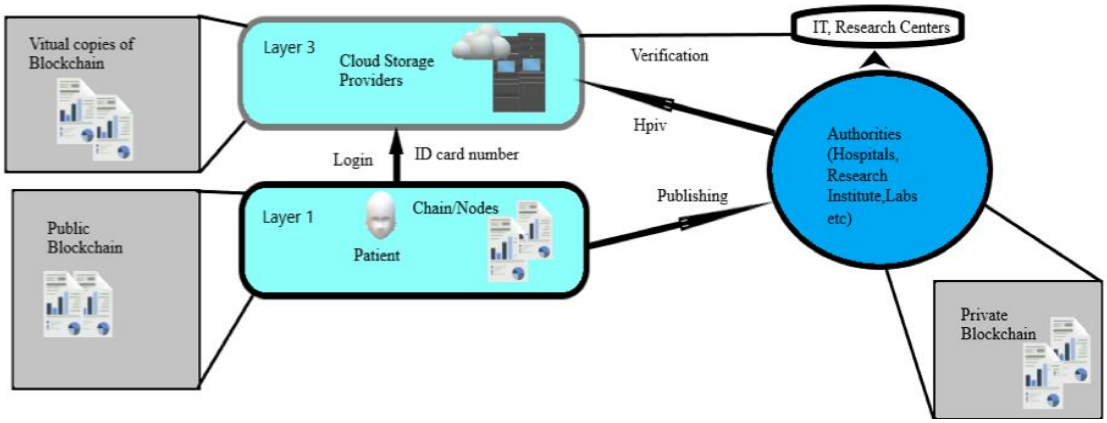


Figure 1: The generic model integrating the eHealth for reforms process in a society (Adopted from [20])

NTP vision is to create an infrastructure and clear a path for achieving the objectives of the vision 2030. NTP consists of eight directions, which are termed as "themes". Among the eight themes, one of the most important ones is the health care services.

The importance the KSA government gives to health is clear from the NTP 2020 document [5]. In NTP, the health theme is the first and given top priority. In fact, the health restructuring is considered a key element to achieve a vibrant society. The NTP aims to promote a new model of health care that has a focus on disease prevention, and then improving health awareness. The NTP has key parameters to improve access to health. For this, NTP advocates the optimal coverage, equitable geographical health facilities distribution, and health facilities expansion through e-health digital solutions.

3.1. NTP health objectives

The NTP health objectives which are directly or indirectly in line with the Euro Health 2020 are as follows:

1. Access facilitation to Health Services

2. Health services quality and efficiency improvement

The health theme of the NTP 2020 and Euro Health 2020 has a number of similarities. The reason is that Euro Health 2020 sets standards and benchmarks for all the world countries and societies. Many countries take help from Euro 2020 for improving the health system in their countries. The major differences are in implementation techniques and methods for the realization of the key points. In a comparative study, it was discovered that all the points of NTP are in line with Euro Health directly or indirectly.

4. Smart Health Framework: Multi-Layer Blockchain

The current interoperability challenges in the health care sector can be solved with the help of blockchain. Blockchain allows individuals, health care providers, and researchers to share health data in a secure way. In this paper, an e-Health system architecture is proposed that will not only improve the security of data but also facilitate the IT sector and research community. A blockchain-based access-control manager is proposed to keep store health records that would share a nationwide interoperability roadmap. Interoperability is a cardinal factor of any system that deals with the Precision Medicine Initiative (PMI) and the Patient-Centered Outcomes Research (PCOR).

Blockchain-based IT infrastructure has come forward to increases the potential development of precision medicine and advance medical research. Blockchain is digitized and decentralized technology that forms a shared database to store data. It has three main components: a distributed network, a shared ledger, and digital transactions. The first layer consists of participant nodes that store an identical copy and helps collectively towards certifying and validating digital transactions for the network. The second layer allows records to be stored. The members who run the algorithm to add new ones verify these transactions. If the majority of participants agree with the added transaction, it approves the validity of that transaction. No single member can permit or alter the data once it has been added.

A blockchain is a digitized, decentralized, distributed public ledger that acts as a shared and synchronized database that records cryptocurrency transactions. While blockchains are essentially decentralized databases, there is no primary ownership of the data [22, 23]. Through collaboration, users decide which data are added to the blockchain while ensuring that identical copies of the data are received and automatically updated [24].

Blockchain removes the need for a third-party and saves the financial mechanism. Finally, blockchain technology can permit the world's poorest person to access a unilateral and a corrupt-free system of wealth [25].

The digitized data is authenticated and secured with the encryption method [26]. Blocks are stored in a linear, chronological order. There are currently no open standards or implementations of blockchain that utilize this approach, but the research supports the feasibility of the proposed solution.

The proposed method has three layers:

1. The constraint nodes (IoT, Sensors, Gateways, Patients)
2. Authorities (Hospitals, Research Institute, Labs, etc.)
3. EHR layer (Cloud Providers, Storage Banks)

A multi-layer framework based on blockchain in IoT infrastructure has been presented. The Elliptical curve approach for Cryptography was used instead of a simple technique to make it more secure and reliable. The authentication and key generation processes moved from the Internet of things (IoT) to their respective gateways resulting in reducing the computational burden. Patient virtual blocks are introduced in the proposed model that will hide the original identity of the patient to make it more secure and efficient.

The authors used *MIRACL* which is a C library that allows secure communication in the blockchain. Multi-precision integers, datatypes, and cryptography are implemented through this library. Any blockchain for health care needs to be public that facilitates the research centers to promote research of the medical domain. The blockchain data contains health records, images, reports, medications, etc. An index is proposed for the information stored in the blockchain, which is like the card. That card has the metadata that has the exact location of the data of respective patients. The transactions are done in encrypted blocks. The health blockchain stores all the data related to medical history. Various metadata handle all the frequently used queries. The data can be retrieved from the hospital entities, laboratories, sensors, or from the patient itself.

All the medical data against each patient will be stored in the virtual blockchain passes towards research centers. It is an important feature of the proposed model. It includes many applications including data mining, text classification, text mining, real-time analysis, batch queries, machine learning [27]. It would be a beneficial tool for the research community and doctors. It would be helpful in finding optimal solutions based on genetics makers and real-time analysis. It will connect the whole community into one platform. The important queries can be answered through a proper and authentic channel.

The following scenarios for the proposed model as shown in Figure 1 are proposed:

1. Data has been collected by the patient from chain/nodes to create EHR.
2. EHR has been accessed by the patient or provider (authority).
3. Now, the patient can add more blocks due to his visit to the provider as seen in Figure 2.

Figure 2 shows the first layer uses public blockchain technology as it covers the new entry into different blocks. These new entries can be from patients, hospitals and sensors. They have to complete their initial registration. The block parameters include patient ID, ID card number, services, the hash of the earlier block. After the registration process, sensors can be accessed through private keys. The user itself has the public key used to create a digitized signature to accept the access permission. A digital signature is imposed for the verification of documents including lab tests, MRIs and prescriptions, etc.

The second layer includes the private blockchain infrastructure. The health record of patients is created by healthcare institutes. Let's say, a patient visited any hospital, the permitted providers created the block in the system. The data will be into the cloud storage provider for the presence of a unified record. The provider has to restrict the right to avoid EHRs abuse. Every patient holds a single block when a visit is made to any hospital containing the information of time and date. The next time when he visits any other clinic, the data is being added to any other chain/node. On the cloud side, they are attached in a linear order to maintain history and current progress. This helps in saving time.

Moreover, the patient would have the authority to choose a virtual identity. He/She can use random value M as a public key by incorporating public key as $H_{\text{pub}} = M.H$ and private key as $H_{\text{priv}} = M.H.S$. In the second layer, the patient is authenticated by using its private key and then the cloud service provider decrypting the data by using its public key. That's how identity can be secured. Every individual owns his chain block in the blockchain, triggered by other health and IT sectors. The encrypted data is further passed for processing and storage. A notification is sent to the patient when new data is added in the blockchain. Similarly, the patient itself can store data into blocks of blockchain with digital signatures for authorization purposes.

The user enjoys full access to his data and control of sharing his data. There would be a set of permissions for who can write and update data in his block. The permissions would be flexible

instead of "all or none". The user can choose the control of his block, who can write data, time period of excess and types of data that can be shared. The user can change permission once they are chosen. These steps allow transparency in society.

When a health care provider gets the permission of access from the user, he can query the blockchain and use the ID for authentication. They can use a customized application to analyze the data. That is how a user is a single entity having control over his data and the power to grant permission for collaboration and communication during disease treatment. Digitally signed transactions combined with decentralized nature makes blockchain a secure network by having control over the network's resources.

Blockchain offers many benefits to the IT health sector. It is open-source software, developed by experts. It is transparent, reliable, responds faster against speedy changes that couldn't be chased by using closed and patented software. Open-source software offers wide innovation and the possibility of choosing options according to needs and demands. Blockchain technology runs on commodity hardware, provides low-cost high computation. They built by various vendors based on open standards. They are based on the best practices in the market because they are easy to implement and discard the complex point to point data integration. Patients, health care entities, the research community can share their accurate data in a timely manner. This

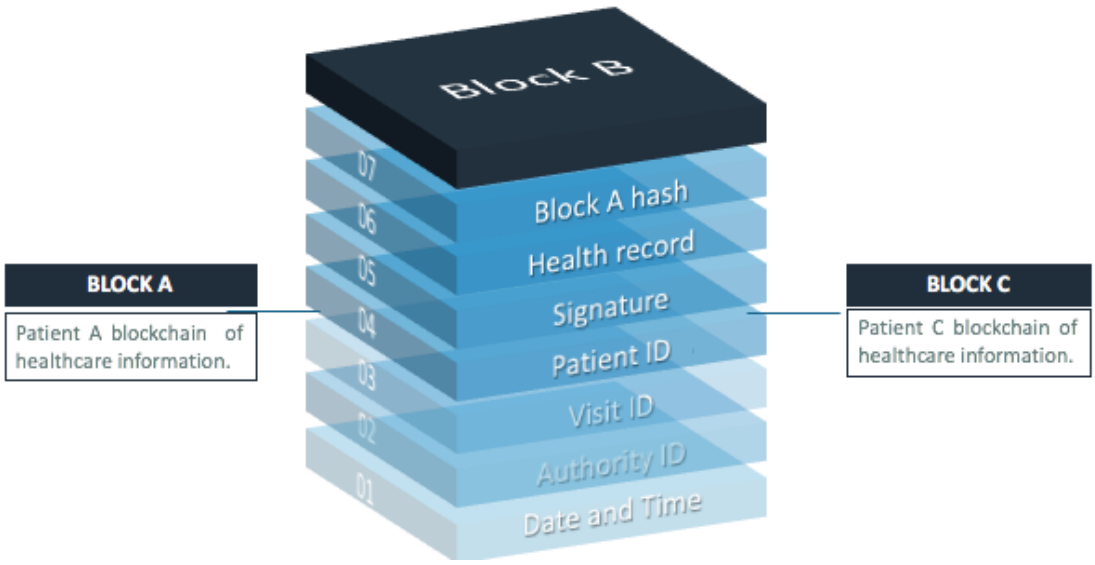


Figure 2: Patient health records in the blockchain. Each patient has one blockchain of healthcare alone

technology has been accepted and analyzed as a secure and efficient system across many private and government agencies. However, comparing cloud-based solutions with blockchain, the blockchain has gained many benefits with regards to security, flexibility, cost, and reliability. Thus, the cost and security of blockchain technology seem to be a much better solution to cloud cost and security concerns. With these solutions, for example, no third parties can access data due to each node only stores encrypted medical records so that providers and patients can help keep their own information safe from prying eyes.

The developed framework contains three layers as described earlier. Taking advantage of the blockchain technology with the cloud computing offered by [7], the proposed framework could store and share seamless health data between different entities, reducing the cost of securing and protecting patients' records. Additionally, the privacy of patients and security in the healthcare domain is preserved in order to maintain the immutability of EHRs. Thus, combining blockchain with the approach of [7] provides better performance with cryptographic adoption to make patient data secure. Moreover, the proposed framework offers immutable medical records to be an essential part of the blockchain mechanism.

5. Discussion

eHealth is an umbrella for connecting patients, doctors, research community and government agencies. In the context of KSA, power is ruled by monarchs. Reform is an ongoing process to improve socio-economic and political values. The government is also trying to improve the health sector through marketization and privatization. This study presents the impact of the health sector in any state, problems towards previous models and government policies of KSA. Additionally, KSA 2020 vision for the health sector is also discussed. eHealth technology benefits people in a more efficient way and it is more cost-effective.

The proposed model use interoperability objectives based on open standards to establish national technology infrastructure. An important tool to connect healthcare centers with the IT sector, the research community, and various agencies has been introduced. The purpose behind this is to improve the health and lifestyle of people in KSA. The model offers services towards analyzing the statistics and generating a report. The excess hardware capacity of multi-layer blockchain could be helpful for resource allocation, privacy, and reliability resulting in a faster solution and processing time. Multiple incentives and benefits will promote hospitals to develop solutions that allow for patient's data to be reachable across different hospitals, and not just within a single health system and hospital.

The utilization of the proposed method has enough potential to connect millions of users, health care entities, health care providers and medical researchers to share vast amounts of health data, genetics, lifestyle with guaranteed privacy and security.

6. Use cases and analysis of current practices

Increasing pressures and demands on health systems call for a significant change in the way of organizing and managing the delivery of health services. The quality of service expected from the Primary Health Centers (PHCs) and hospitals along with an ever-increasing number of patients in KSA hospitals calls for a productive change in the procedures of managing and delivering health services. Various doctors working in government hospitals like King Saud Hospital Unayzah, Qassim University Clinics, King Fahad Specialist Hospital Buraidah and Dr. Sulaiman Al-Habib hospital, Buraydah in the private sector were consulted in order to analyze the current practices in these health centers.

These include but are not limited to the digitization of care services. It was found that almost all of the hospitals have digital patient health records. The patients are allotted an identification number during their first visit to the facility. This ID is then used

to manage all consultations, laboratory tests, provision of medicines, etc. This enables these centers to provide a better quality of service to the patients. Furthermore, it reduces a lot of paperwork, though it is still being maintained. However, Sulaiman Al-Habib hospital enables the patients to register online. Furthermore, the patients can book and manage appointments online besides having access to their laboratory and radiology results.

- Sulaiman Al-Habib, being the only private hospital discussed in this research, could be considered as one of the patient-centered hospitals. Everything revolves around the patients and the patients have the freedom to choose the doctors for their appointments. Moreover, patients can ask questions to the visited doctors and give ratings to the doctors as well.
- The influx of patients to secondary and tertiary care hospitals has been increased over the last few years. This pattern is more obvious in the case of Emergency and adversely affects the provision of timely care to critical patients. Therefore, many patients who are not in need of emergency treatment are being referred to the PHCs. Moreover, doctors serving in the hospitals are being asked to help the PHCs so as to manage the increased load. This has resulted in better management of seriously ill patients.
- Although most of the hospitals in KSA have digitized their records, yet there is no process to share the patients' data between different health centers. King Faisal Specialist Hospital, Riyadh has deployed a solution by Informatica which is HL7 compliant. The following are some of the common issues faced by the patients:
- The patients of government hospitals have got no access to their medical records. In the case of private centers, the data is accessible though not shared with other hospitals or pharmacies. In many cases, if a patient visits a different center, the medical tests have to be reconducted. Sometimes, patients are given CDs containing their Magnetic Resonance Imaging (MRI), X-Ray results. Even in this case, the CDs can malfunction and the patient has to revisit the medical facility.

Conclusion and Future Works

In this paper, a multi-level blockchain eHealth system is proposed to provide seamless electronic health records to patients. A detailed comparison between Saudi Arabia's National Transformation Program (NTP) and the Euro Health 2020 plan is made. It was observed that the main difference lies in the implementation techniques and methods employed for the realization of these plans. The proposed blockchain-based model is composed of constraints nodes, authorities and cloud providers. The user has complete control of his data. Moreover, the current practices being employed in Qassim, KSA has also been analyzed. It was found that the private hospitals give access to medical reports of their patients besides allowing them to manage their appointments. On the other hand, access to the government hospitals' medical records is very limited. In the future, the study can be extended to different regions of KSA.

Acknowledgments

The authors gratefully acknowledge Qassim University, represented by the Deanship of Scientific Research, on the material support for this research under the number 3578-coc-2018-1-14-S during the academic year 1439 AH / 2018 AD.

References

[1] K. A. Stroetmann, "Analysis and Typology of Global eHealth Platforms – A Survey on Five Continents" Stud Health Technol Inform, 209, 162–169, 2015. https://doi:10.3233/978-1-61499-505-0-162

[2] D. Langbecker, L. J. Caffery, N. Gillespie, A. C. Smith, (2017). "Using survey methods in telehealth research: A practical guide" Journal of Telemedicine and Telecare, 23(9), 770–779, 2017. https://doi.org/10.1177/1357633X17721814

[3] E. E. Ali, L. Chew, K. Y.-L.Yap, "Evolution and current status of mhealth research: a systematic review" BMJ Innovations, 2(1), 33-40, 2016.

[4] S. A. Talesh, "Data breach, privacy, and cyber insurance: How Insurance Companies Act as "Compliance Managers" for Businesses" Law & Social Inquiry, 43(2), 417–440, 2017. https://doi.org/10.1111/lsi.12303

[5] KSA Ministry (2018). "National Transformation Program | Saudi Vision 2030". Retrieved from https://vision2030.gov.sa/sites/default/files/attachments/NTP%20English%20Public%20Document_2810.pdf.

[6] C. Harrison, I. A. Donnelly, "A theory of smart cities" in 55th Annual Meeting of the International Society for the Systems Sciences (ISSS), Hull, 2011.

[7] H. F. El-Sofany, K. F. Al-Otaibi, M. S. Alsanea, "Data exchange model of patient records in the Kingdom of Saudi Arabia using cloud computing" International Journal of Innovation, Management, and Technology, 3(5), 580–589, 2012. https://doi.org/10.7763/IJIMT.2012.V3.301

[8] L. Zhuhadar, E. Thrasher, S. Marklin, P. O. Pablos, "The next wave of innovation - Review of smart cities intelligent operation systems" Computers in Human Behavior, 66 (2017) 273–281, 2017 https://doi.org/10.1016/j.chb.2016.09.030

[9] A. Bhowmick, E. Francellino, L. Glehn, R. Loredo, P. Nesbitt, S. W. Yu, "IBM intelligent operations center for smarter cities". IBM International Technical Support Organization.

[10] Kylie Wansink "Saudi Arabia - Telecoms, Mobile and Broadband - Statistics and Analyses". Retrieved from https://www.budde.com.au/Research/Saudi-Arabia-Telecoms-Mobile-and-Broadband-Statistics-and-Analyses (6 May 2019)

[11] D. Gooch, A. Wolff, G. Kortuem, R. Brown, "Reimagining the role of citizens in Smart City projects" in ACM International Joint Conference on Pervasive and Ubiquitous Computing and ACM International Symposium on Wearable Computers, Osaka, Japan, 2015. https://doi.org/10.1145/2800835.2801622

[12] H. Chourabi, T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, T. A. Pardo, H. J. Scholl, "Understanding Smart Cities: An Integrative Framework" in 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 2012. https://doi.org/10.1109/HICSS.2012.615

[13] D. H. Gustafson, J. C. Wyatt, "Evaluation of eHealth systems and services" BMJ, 328, 1150, 2014. https://doi.org/10.1136/bmj.328.7449.1150

[14] D. J. Cook, G. Sprint, G. E. Duncan, R. Fritz, "Using Smart City Technology to Make Healthcare Smarter" Proceedings of the IEEE, 106(4), 708–722, 2018. https://doi.org/10.1109/JPROC.2017.2787688

[15] Regional officer for EU, "Health 2020: the European policy for health and well-being," World Health Organization, Sep. 2019.

[16] C. H. Lee, K.-H. Kim, "Implementation of IoT system using block chain with authentication and data protection" in IEEE International Conference on Information Networking (ICOIN), Chiang Mai, Thailand, 2018. https://doi.org/10.1109/ICOIN.2018.8343261

[17] C. O. Rolim, F. L. Koch, C. B. Westphall, J. Werner, A. Fracalossi, G. S. Salvador, "A Cloud Computing Solution for Patient's Data Collection in Health Care Institutions," in IEEE Second International Conference on

eHealth, Telemedicine, and Social Medicine, St. Maarten, Netherlands Antilles, 2010. https://doi.org/10.1109/eTELEMED.2010.19

[18] A. F. Hussein, N. ArunKumar, G. Ramirez-Gonzalez, E. Abdulhay, J. M. R. S. Tavares, V. H. C. de Albuquerque, "A medical records managing and securing blockchain based system supported by a Genetic Algorithm and Discrete Wavelet Transform" Cognitive Systems Research, 52, 1–11, 2018. https://doi.org/10.1016/j.cogsys.2018.05.004

[19] M. Moran, "Understanding the welfare state: the case of health care" The British Journal of Politics and International Relations, 2(2): 135–160, 2000. https://doi.org/10.1111/1467-856X.00031

[20] T. Dey, S. Jaiswal, S. Sunderkrishnan, N. Katre, "HealthSense: A medical use case of Internet of Things and blockchain" in IEEE International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 2017. https://doi.org/10.1109/ISS1.2017.8389459

[21] O. Alphand, M. Amoretti, T. Claeys, S. Dall 'Asta, A. Duda, G. Ferrari, F. Rousseau, B. Tourancheau, L. Veltri, F. Zanichelli, "IoTChain: A blockchain security architecture for the Internet of Things" in IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 2018.
https://doi.org/10.1109/wcnc.2018.8377385

[22] V. A. Patel, "A framework for secure and decentralized sharing of medical imaging data via blockchain consensus" Health Informatics J., 25(4), 1398–1411, 2018.
https://doi.org/10.1177/1460458218769699

[23] M. Conoscenti, A. Vetrò, and J. C. D Martin, "Blockchain for the Internet of Things: A systematic literature review" in IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), Agadir, Morocco, 2017.
https://doi.org/10.1109/AICCSA.2016.7945805

[24] W. J. Gordon, C. Catalini, "Blockchain Technology for Healthcare: Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability" Computational and Structural Biotechnology Journal, 16, 224–230, 2018.
https://doi.org/10.1016/j.csbj.2018.06.003

[25] H. Kaur, M. A. Alam, R. Jameel, A. K. Mourya, V. Chang, "A proposed solution and future direction for blockchain-based heterogeneous medicare data in cloud environment" Journal of Medical Systems, 42(8), 156, 2018. https://doi.org/10.1007/s10916-018-1007-5

[26] T. K. Mackey, G. Nayyar, "A review of existing and emerging digital technologies to combat the global trade in fake medicines" Expert Opinion on Drug Safety, 16(5), 587–602, 2017.
https://doi.org/10.1080/14740338.2017.1313227

[27] Albahli, Saleh. "A Deep Ensemble Learning Method for Effort-Aware Just-In-Time Defect Prediction." Future Internet 11, no. 12 (2019): 246.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 48-57 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Windowing Accuracy Evaluation for PSLR Enhancement of SAR Image Recovery

Mohamed Bakry El_Mashade[*,1], Haithem Akah[2], Shimaa Abd El-Monem[2]

[1]*Electrical Engineering Dept., Faculty of Engineering, Al_Azhar University, Nasr City, Cairo, Egypt*

[2]*Space Communication Dept., National Authority for Remote Sensing and Space Sciences, Nozha El-Gedida, Cairo, Egypt*

ARTICLE INFO

Article history:
Received: 19 October, 2019
Accepted: 28 December, 2019
Online: 15 January, 2020

Keywords:
Synthetic Aperture Radar
SAR image processing
Windowing
SAR image quality
Range-Doppler Algorithm

ABSTRACT

Synthetic aperture radar (SAR) is an imaging device mounted on a moving platform. Its ability to identify a weak target from a nearby strong one depends upon the peak side lobe ratio (PSLR). This paper is intended to ameliorate such important ratio through the use of windowing of the transmitted pulse and studying the noise sensitivity of the resulted image. Two experiments are simulated for different types of windows using point target and earth images. PSLR, dynamic range, spatial resolution, radiometric resolution, mean and standard deviation are evaluated for each window of the point target image. Also, the sensitivity to noise is discussed through the calculation of the mean square error, root mean square error, signal-to-noise ratio (SNR), peak SNR for the earth image. Experimental results show that Triangular, Gaussian and Hann achieve the highest PSLR performance with different reductions in their resolution, whilst Taylor has the worst performance due to its high sensitivity to noise. The remaining ones give intermediate performance.

1. Introduction

Synthetic aperture radar (SAR) can be viewed as an aerial remote surveillance tool because of its ability to originate 2D and 3D high-resolution mapping with all-weather visibility. It is an imaging radar mounted on a moving platform. To originate an image of a target, it sequentially transmits electromagnetic waves and collects the backscattered echoes through the radar antenna. It employs the motion of the radar antenna over a target region to provide finer spatial resolution. In other words, a target scene is illuminated by successive pulses of radio waves with wavelengths of a meter down to several millimeters and the echo of each pulse is received and recorded. The consecutive times of transmission/reception are translated into different locations due to the platform movement. By processing these successive echoes, the recordings from multiple antenna positions are combined to establish an image for the target under test. Typically, an appropriate coherent combination of the received signals allows the construction of a virtual aperture which is much longer than the physical antenna length. In this regard, the larger the aperture, the higher the image resolution will be, regardless of whether the aperture is physical (a large antenna) or synthetic (a moving antenna). As a consequence of this, high-resolution images, with comparatively small physical antennas, can be created. In general, SAR employs the different locations of the sensor, as it moves along the flight path, to simulate a large antenna from a smaller one. This enables its sensor to provide high resolution imagery which will not be degraded with distance as in the case of traditional radar systems with large antennas. For these reasons, SAR becomes a more popular imaging sensor in most practical applications [1-5].

There are different modes of operation of SAR systems. Since the strip-map SAR mode is the most interesting one of these modes, it is chosen to be the scope of treatment in this paper. The appellation of this mode originates from the fact that it is associated with the type of SAR that can continually map strips of the ground as the aircraft flies by. In this mode, the electromagnetic energy is transmitted and its associated radar beam is used to illuminate a specified area that is called footprint. The radar antenna is pointed along a fixed direction according to the flight path, and the antenna footprint covers a strip on the illuminated surface of a point target [6-8]. The length of the strip is limited by the distance to which the sensor moves through the travel. In this regard, SAR strip-map data acquisition geometry is illustrated in Figure 1. In this figure, the flight direction is called azimuth direction, whilst the perpendicular to which is called range direction; the length of which extends from the radar antenna to the target [9,10].

* Mohamed Bakry El_Mashade Email: Mohamed.B.ElMashade@azhar.edu.eg

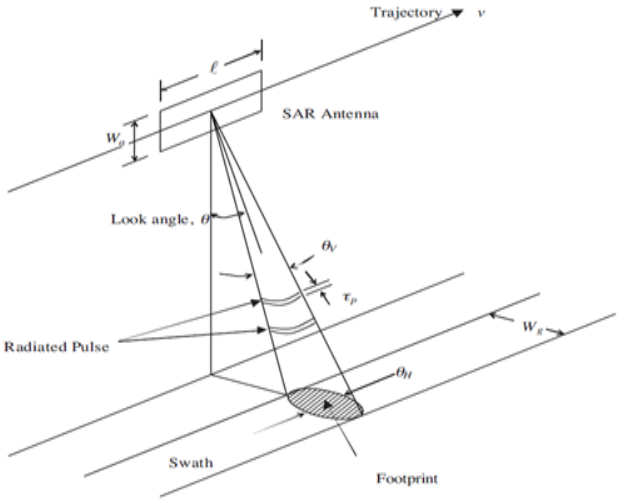


Figure 1: SAR data acquisition geometry

The received SAR signal can be considered as a two-dimensional (range and azimuth) signal; the construction of which is as shown in Figure 2.

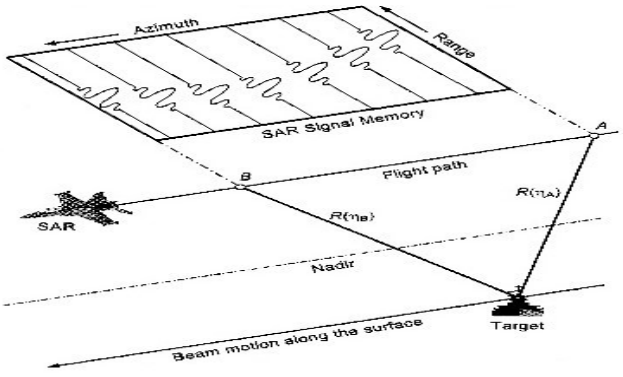


Figure 2: SAR data construction as a 2-dimensional signal

In this figure, it is shown that when the sensor is at position A, the target lies in a radar beam. So, the received signal is recorded in one row of the memory. While the sensor moves, it sends more pulses and the received signal is written in another memory location. When the sensor is at position B, on the other hand, the target leaves the beam and the last received signal is preserved. This two-dimensional signal is called "raw data", "phase history data", or "echo". Its coordinates are the range (fast) time and the azimuth (slow) time. Power spreading of this raw data represents the input to the SAR signal processing/image formation algorithm; the main function of which is to compress the signal energy in order to reconstruct the final image [11-13].

The final version of image processing is called single look complex (SLC) image. To extract the SLC image from the raw (echo) data, there are several algorithms including range-Doppler algorithm (RDA), chirp scaling algorithm, SPECAN algorithm, and omega-k algorithm. Due to its efficiency, accuracy, maturity, and ease of implementation, RDA is the most widely used algorithm. For this reason, it is chosen to be an image formation processor in this research. This algorithm requires some information about the transmitted pulse and imaging geometry, such as range and satellite velocity, to construct the range and azimuth compressions, respectively. Its idea is based on matched filtering process to perform such types of compression. Additionally, it uses block processing to achieve frequency domain representation of both range and azimuth separately. Accordingly, RDA has some steps, as Figure 3 demonstrates, that must be followed in order to formulate the SLC image. In these steps, the FFT-based measurement represents the error resulting from the effect known as leakage. It occurs when calculating a non-periodic data by FFT. To overcome such error, the window function must be applied [14-16]. Thus, by applying a window function to the algorithm shown in Figure 3-a, a new modified version of this algorithm is clarified in Figure 3-b.

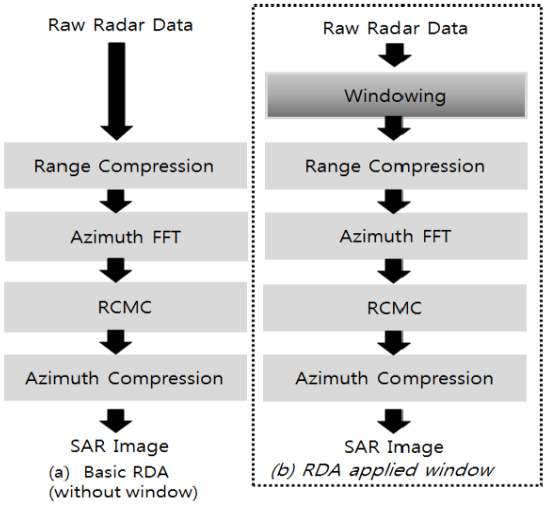


Figure 3: (a) Block diagram of basic RDA (b) Block diagram of RDA applied window

It is of interest here to enhance the peak side lobe ratio (PSLR) of SAR image. This can be achieved by applying a window function to the transmitted pulse. Gaussian, Kaiser, Taylor, Chebyshev, Triangular, Hann, Blackman, Flattop and Tukey window functions are used to generate different raw data which is applied to RDA to get the final SAR image for each window. Additionally, PSLR, dynamic range, spatial and radiometric resolutions are studied. Moreover, the mean square error (MSE), root means square error (RMSE), signal-to-noise ratio (SNR) and peak SNR (PSNR) are computed for the above mentioned window functions.

The paper is structured as follows: section 2 is concerned with the model description of the underlined point of research. Section 3 discusses the problem under examination. SAR raw data generation with different window functions is outlined and a comparison between the accuracy of the final images is made in section 4. In section 5, the windowing range reference signal as well as the result of RDA applied window is presented and the errors in the final images are compared through the metric parameters. Finally, section 6 summarizes our concluded remarks.

2. Model Description

In SAR system, the image formation processor produces an image that is a 2-dimensional mapping of the illuminated scene. The formed image is interpreted in the dimensions of range and cross-range or azimuth. The range resolution of a SAR image is directly related to the bandwidth of the transmitted signal and the cross-range is inversely proportional to the length of the antenna aperture. Therefore, high range resolution is achieved by transmitting wide bandwidth waveforms, and high cross range resolution is achieved by coherently processing returns received from a variety of positions along a flight path to emulate a large aperture. Accordingly, pulse compression is used to maintain the average transmitted power of a relatively long pulse while obtaining the range resolution corresponds to a short pulse. In this regard, linear frequency modulation (LFM) is one of the most common pulse compression techniques. In this type of modulation, the transmitted signal has a form given by:

$$S(t) = W_r(t)\cos(2\pi f_0 t + \pi k_r t^2) \quad (1)$$

In the above expression, t represents the fast time, f_0 denotes the carrier frequency, k_r is the range chirp or FM rate, and $W_r(t)$ is the pulse envelope of the transmitted signal.

The received signal is a replica of the transmitted one with the exception that it is time delayed, attenuated, phase-shifted, and accompanied with additive white Gaussian noise (AWGN). To remove the high-frequency carrier and convert the signal to the baseband, the quadratic demodulation is used and the result can be expressed as:

$$s(t,\eta) = \sum_{\ell=1}^{M-1}\left\{\chi_\ell W_r\left(t-\frac{2R_\ell(\eta)}{c}\right)w_a(\eta-\eta_c)e^{-j4\pi\left(\frac{f_0R_\ell(\eta)}{c}\right)} e^{j\pi k_r\left(t-\frac{2R_\ell(\eta)}{c}\right)^2}\right\} + n_\ell(t,\eta) \quad (2)$$

In this formula, t denotes the range time, η symbolizes the azimuth time, χ represents the attenuation factor from reflection at the target, $n(t, \eta)$ denotes an additive white Gaussian noise, $w_a(\eta-\eta_c)$ is the azimuth beam pattern amplitude modification, which is illustrated in Figure 4, and $2R(\eta)/c$ refers to the time delay.

Eq.(2) represents a summation of the reflections from M different point targets. This formula is used in MATLAB to generate all the reflections from the area that is required to be imaged during the flight direction. This complex style equation constitutes the signal in its raw form (echo). It is a function of both range domain (fast time) and azimuth domain (slow time) where the energy is spreaded in them.

Basic RDA uses a rectangular window in its transmission. Through the range compression process, which is performed in the frequency domain using the correlation between the raw data and a replica of the rectangular transmitted pulse, Eq.(1) can be expressed as:

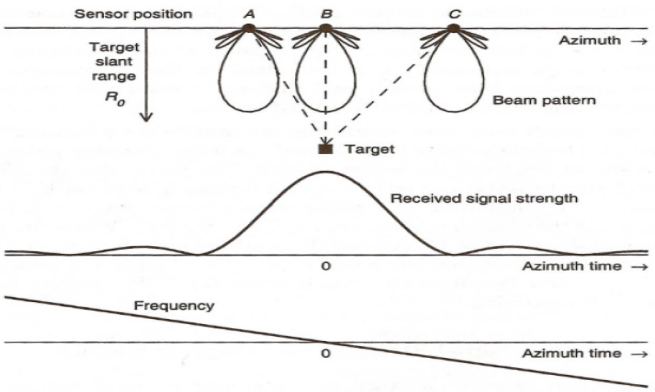


Figure 4: Azimuth Beam Pattern

$$\overline{S}(t) = \cos(2\pi f_0 t + \pi k_r t^2) \quad (3)$$

The result of this correlation is transformed back to the time domain to completely perform the range compression process. On the other hand, to prepare the data for azimuth compression, the second step is to carry out azimuth FFT which is needed to transform the data into the range-Doppler domain. In this domain, RCMC can be applied due to its hyperbolic trend with respect to the azimuth time. Then, azimuth compression is constructed. The result of this compression is transformed back to the time domain using inverse fast Fourier transform (IFFT) to obtain the final image [12].

Due to the importance of windowing in enhancement of PSLR of SAR image generated by RDA, we are going to discuss its basic concepts. In this regard, a window is defined as a symmetrical real function where its weights are applied to the signal spectrum. The maximum value is applied in the middle of the spectrum and roll off towards the edges. The PSLR value for the rectangular window is very high. One of the best methods to reduce it is to apply the smoothing window. Windowing is applied in the frequency domain of the matched filter to reduce the leakage [3]. The effect of windowing appears in smoothing the spectrum where it reduces the discontinuity at the edges. So, the leakage of the main lobe of the resulting compressed pulse is reduced as Figure 5 demonstrates. In contrast, the resolution is degraded due to the reduction of the effective signal bandwidth as a result of compression process [14].

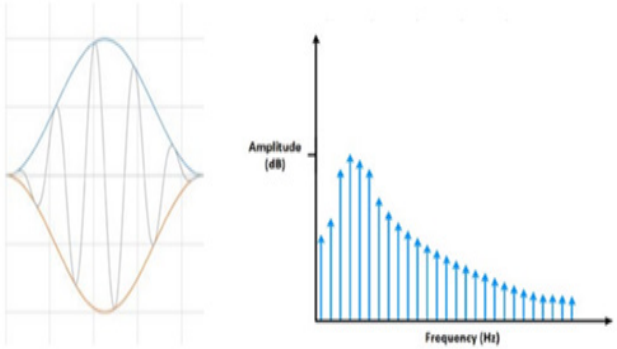


Figure 5: Spectral leakage reduction due to windowing

Rectangular, Gaussian, Kaiser, Taylor, Chebyshev, Triangular, Hann, Blackman, Flattop and Tukey as window functions are used. In our study, two images are simulated to see the effect of each one of these functions on the resulting image. Firstly, the point target image is used for point target evaluation with the aid of impulse response function (IRF). PSLR, dynamic range, spatial resolution, radiometric resolution, and the mean values are computed. Additionally, the sensitivity to channel noise is studied via the calculation of the metric parameters including MSE, RMSE, SNR and PSNR for the second image which is an earth image.

3. Problem Definition

Owing to its ability to measure delicate differences in the surface of the ground, SAR has become one of the most valuable tools for remote sensing of the earth and its environment. In this regard, it has revealed much new information about ground subsidence and the role it can play in natural disasters. One of the key features of the SAR image is target detection, which depends on its ability to differentiate a weak target from a nearby strong one. The PSLR parameter measures the ratio between the largest value of the side lobe to the peak level of the main lobe. For the SAR system IRF, the lower the PSLR is the better the system detectability. Thus, getting the lowest PSLR for any SAR system represents the main goal. The ability of PSLR reduction for SAR image, via windowing of the transmitted pulse, is the objective of this research. The transmitted pulse is changed in accordance with the applied window and the final SAR image will be formulated obeying the specifications of each transmitted signal. Therefore, it is of interest to study the effect of each window on the SAR echo generation and the final SAR image through the image formation processes for the purpose of searching the best one amongst them. Various windows are employed to generate the SAR echo from which the final SAR image, using RDA applied procedure, is obtained. Each window has a different effect on the final image. The difference between the basic RDA and its applied window is illustrated in Figure 3.

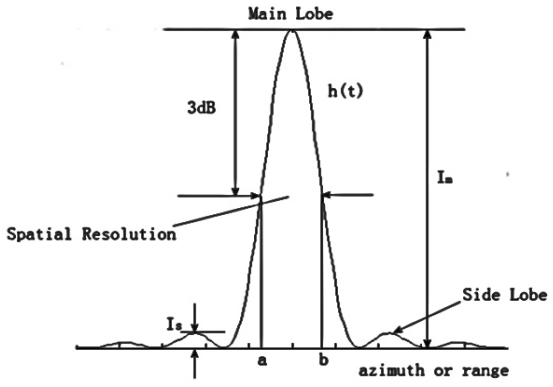


Figure 6: IRF and evaluation of point target parameters

4. Methodology

An experimental approach is implemented by simulating the underlined problem on a PC using MATLAB software. In this simulation, two images are employed. The first one is as bright as possible point target image to make the problem of interest more easily. This image is treated by several windows. Each window is used to generate the raw data with the aid of Eq.(2). This raw data is processed using RDA applied scenario to extract the final SAR image. The SAR image of each window is examined by the IRF. PSLR, dynamic range, spatial resolution, and the radiometric resolution are evaluated in the range direction. These parameters and IRF are displayed in Figure 6. They are calculated with the aid of an interpolation procedure.

The second scene is the earth image. It is used to study the sensitivity of each pulse to channel noise. The mother image is utilized in generating the raw data for several windows using Eq. (2). This raw data is applied to generate the final SAR image using RDA. The noise is added to the raw data to see the sensitivity of each window to such unwanted signal. This noisy raw data is used to generate the final image and MSE along with RMSE, SNR, and PSNR are calculated. In our simulation the following formula is used to generate the noise that is added to the raw data to establish the final data processing:

$$n = \frac{\sigma^2}{\sqrt{2}} \{rand(M,N) + j\ rand(M,N)\} \quad (4)$$

In the above formula, σ^2 denotes the signal power, rand(., .) is the random number generator, M & N represent the number of rows and columns of the image's pixels.

Firstly, the MSE parameter is regarded as the measure of the quality. It is based on a comparison between the generated image and that containing the actual pixel. It has a mathematical form given by:

$$MSE = \frac{1}{M * N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |f(i,j) - g(i,j)|^2 \quad (5)$$

Here, f(i,j) denotes the pixels of the original image, g(i,j) represents the same thing for the generated image, M symbolizes the image's row numbers of pixels, and N refers to the image's column numbers of pixels.

As well, RMSE measures how much error is there between the two images. Mathematically, it is defined as:

$$RMSE = \sqrt{\frac{1}{M * N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |f(i,j) - g(i,j)|^2} \quad (6)$$

Also, SNR is dealing with the signal strength relative to the background noise. This ratio is usually measured in decibels (dB) using the formula:

$$SNR = 10 \log_{10} \left\{ \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [f(i,j)]^2}{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [f(i,j) - g(i,j)]^2} \right\} \quad (7)$$

Additionally, PSNR is an expression for the ratio between the maximum possible power strength of a signal and its accompanied

power of noise. The PSNR parameter is generally expressed in terms of decibel scale. It is mathematically defined as:

$$PSNR = 20\log_{10}\left\{\frac{MAX_F}{\sqrt{MSE}}\right\} \quad (8)$$

In this mathematical expression, MAX_F symbolizes the maximum signal power that is existed in the original image.

Moreover, the mean value, which indicates the level of the reflected signal, is calculated. It is well-known that the higher the mean value is, the better signal level becomes.

In the two images of our experimental simulation, several windows are employed in the generation and focusing processes. They include Gaussian, Kaiser, Taylor, Chebyshev, Triangular, Hann, Flattop, Black-man, and Tukey. These two images are applied to generate the echo signal and consequently a SAR image, of dimension 900 x 1034, is extracted. The needed parameters for the generation and focusing of SAR images are summarized in Table 1. The resulting SAR image has a resolution of 1.5mx1m in range and azimuth, respectively.

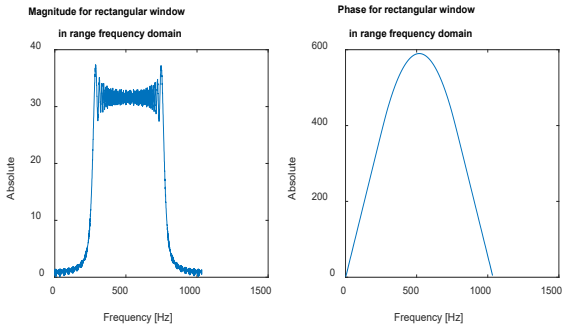
For point target image, the IRF of each window is studied, in range direction, using the character of intensity which is defined as the squared modulus of the complex pixel of the final image. Additionally, normalized PSLR, spatial resolution, radiometric resolution and dynamic resolution are evaluated. For the earth image, on the other hand, MSE, RMSE, SNR, and PSNR are computed for each window as a series of simulations in the case where the noise variance varies from 0 to 3Watt at 16 steps. A total of 1000 samples are taken for each noise variance level in each window.

Table 1: Needed parameters for the generation and focusing of SAR images

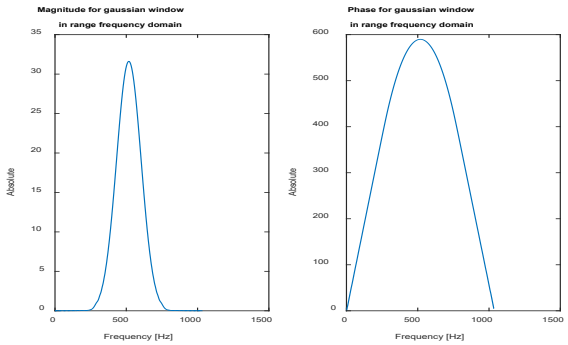
	Parameters		Value	Unit
Parameters for range compression	Range sampling frequency	f_s	4.5	GHz
	Chirp rate	K_r	$4*10^{13}$	s^{-2}
	Chirp length	T_p	2.5	μs
	Range resolution	P_r	1.5	m
Parameters for azimuth compression	Velocity	V	200	m/s
	Wavelength	λ	60	cm
	Range to target at broadside time t = 0	R_0	20	km
	Aperture time	Dur	3	s
	Pulse repetition frequency	PRF	300	Hz
	Azimuth resolution	P_a	1	m

5. Simulation Results

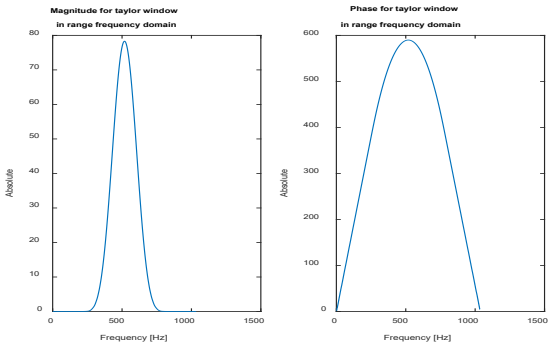
For RDA applied window, the range reference signal is generated as a replica of the transmitted pulse. Accordingly, the windowing range reference signal differs from the rectangular range reference signal which is employed in RDA. Figure 7 depicts the windowing range reference signal for each window in the frequency domain. This signal is used in the application of the two image simulations associated with the point target and the earth. This figure has ten plots each one has two scenes. The first one represents the magnitude while the other refers to the phase. From the displayed results, it is evident that the flattop window in (d) is the narrowest while the rectangular in (a) is the widest. In the two cases, the phase variation has approximately the same behaviour.



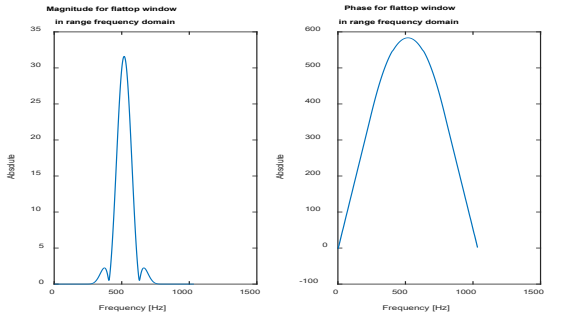
a) Rectangular range reference signal



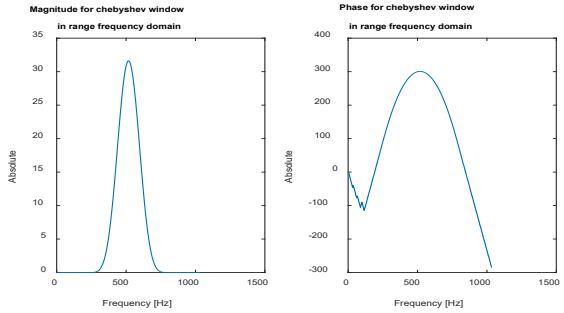
b) Gaussian range reference signal



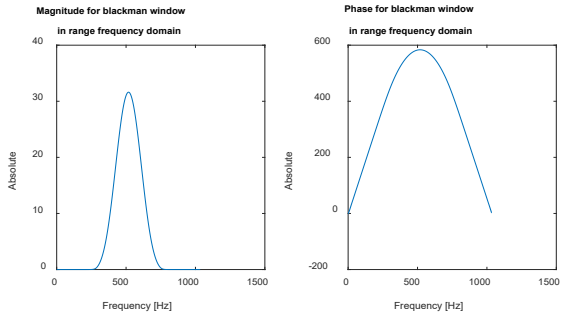
c) Taylor range reference signal



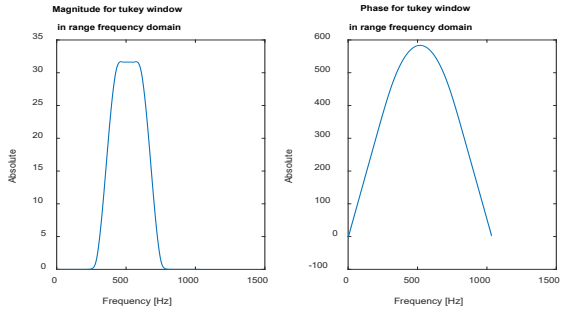
d) Flattop range reference signal



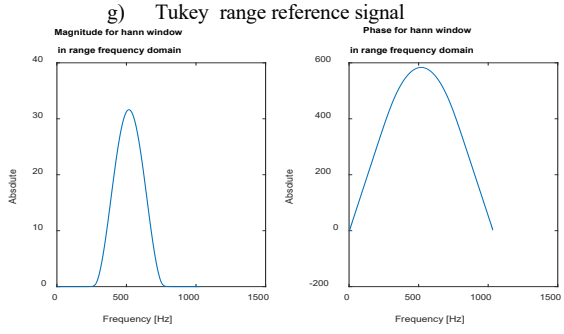
e) Chebyshev range reference signal



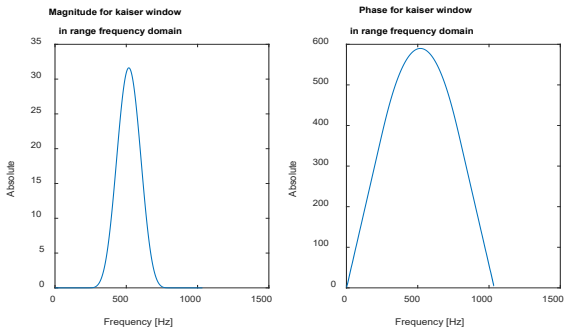
f) Blackman range reference signal



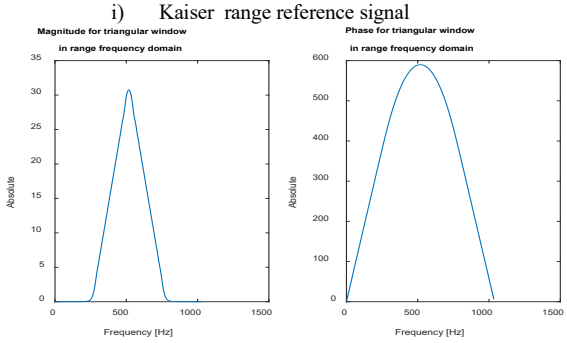
g) Tukey range reference signal



h) Hann range reference signal



i) Kaiser range reference signal



j) Triangular range reference signal

Figure 7: Frequency domain representation of Windowing range reference signals

In the upcoming text, the obtained results of the two experiments are displayed. The point target image is used to study the IRF for each window for the purpose of the range direction evaluation of the point target. The numerical results of this evaluation are summarized in Table 2Table and Table 3.

Table 2: Point target evaluation results

Window	PSLR (dB)	Mean	Dynamic resolution (dB)
	Range		
Rectangular	-14.23	0.0019	211.374
Black man	-20.70	0.0043	196.310
Chebyshev	-20.68	0.0049	190.551
Flattop	-19.94	0.0074	182.486
Gaussian	-20.95	0.0043	196.480
Hann	-19.40	0.0035	192.717
Kaiser	-20.730	0.0046	191.825
Taylor	-20.85	0.0044	211.116
Triangular	-24.89	0.0033	202.516
Tukey	-15.98	0.0030	201.898

Table 3: Resolution of point target evaluation

Window	Radiometric resolution (dB)	Spatial resolution(m) Range (half)
Rectangular	12.985	2.7
Black man	11.420	6.3
Chebyshev	11.156	7.2
Flattop	10.285	10.95
Gaussian	11.377	6.3
Hann	11.802	5.25
Kaiser	11.270	6.75
Taylor	11.334	6.6
Triangular	11.941	4.95
Tukey	12.152	4.53

Let us now turn our attention to the second experiment which is concerned with earth image. The original earth image that is used in our simulation is depicted in Figure 8.

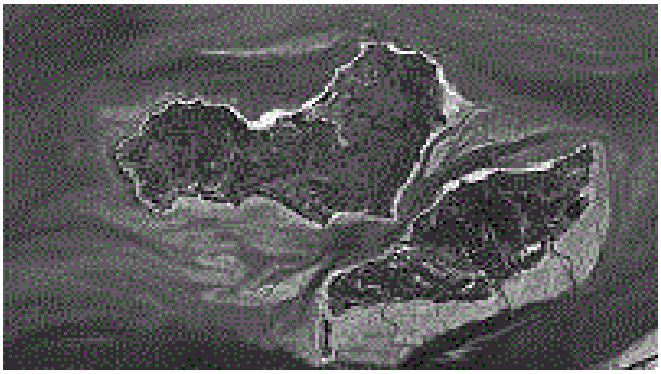
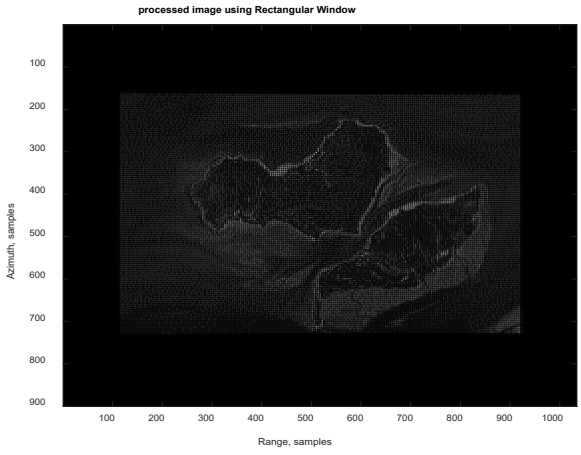
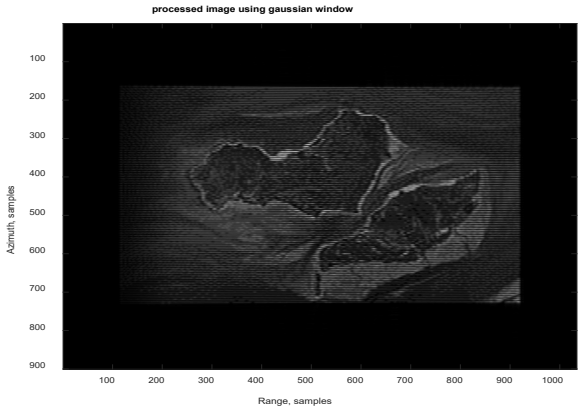


Figure 8: Original earth image

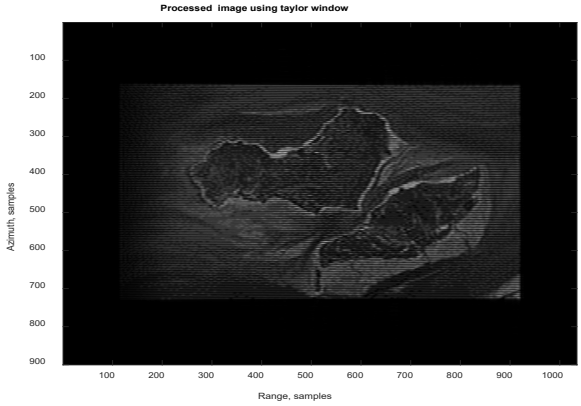
This image is used to generate the raw data from which the final SAR image using RDA algorithm is produced. The extracted image for each window is illustrated in Figure 9. From this figure, it is noted that the final image for the Rectangular window has the best resolution and the Flattop window possesses the worst resolution while the resolution of the rest ones is intermediate between these two indicated windows.



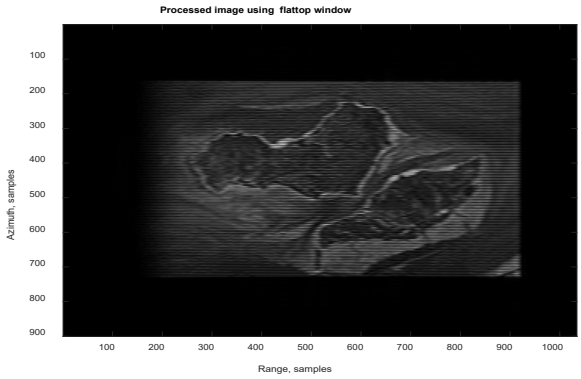
a) SAR image corresponds to Rectangular window



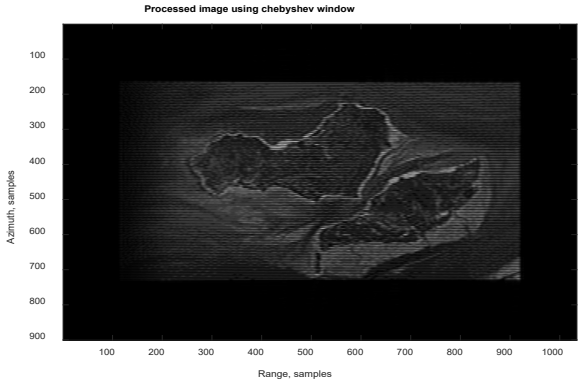
b) SAR image corresponds to Gaussian window



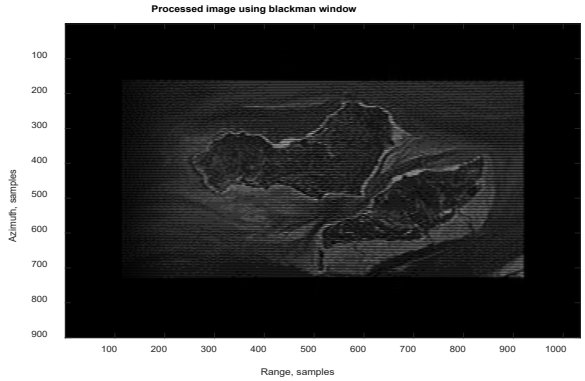
c) SAR image corresponds to Taylor window



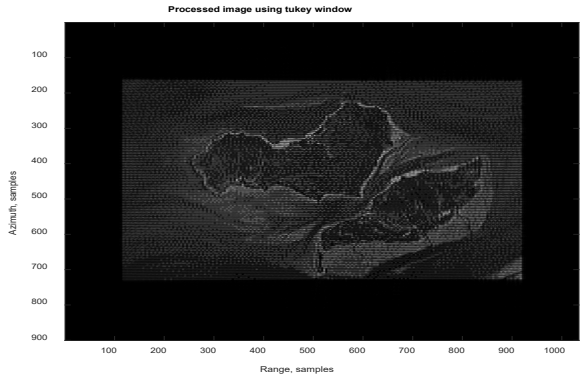
d) SAR image corresponds to Flattop window



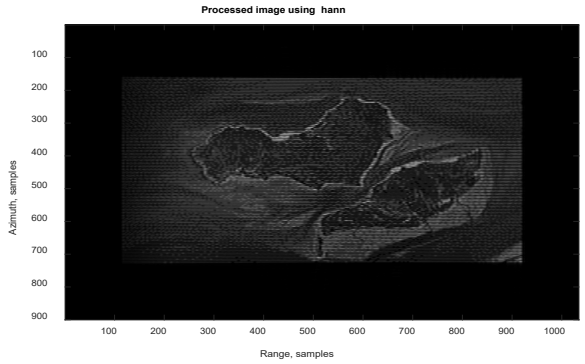
e) SAR image corresponds to Chebyshev window



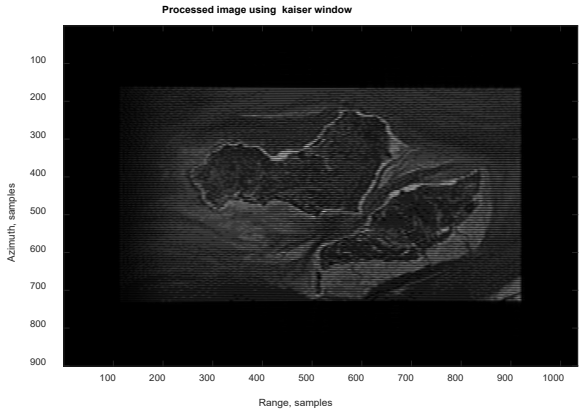
f) SAR image corresponds to Black man window



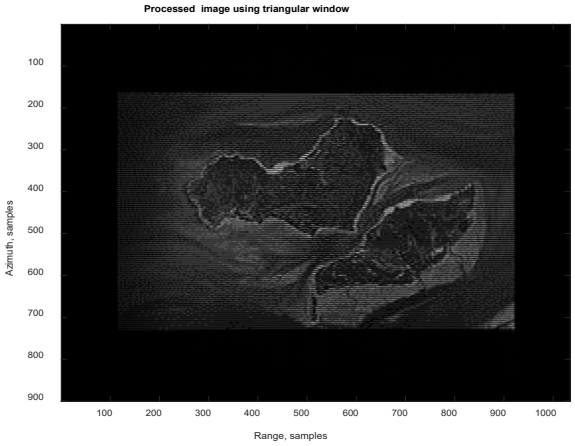
g) SAR image corresponds to Tukey window



h) SAR image corresponds to Hann window



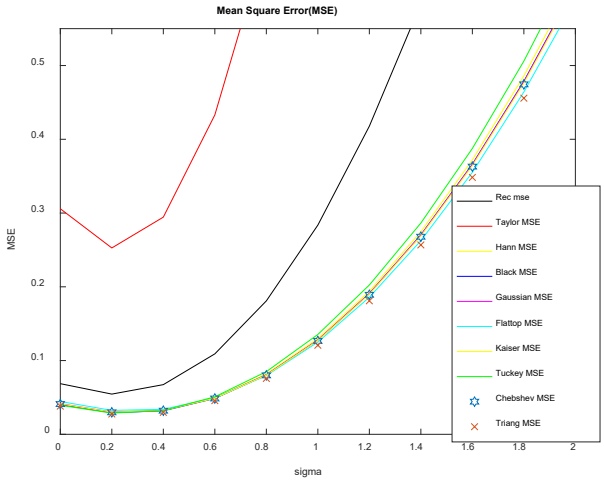
i) SAR image corresponds to Kaiser window



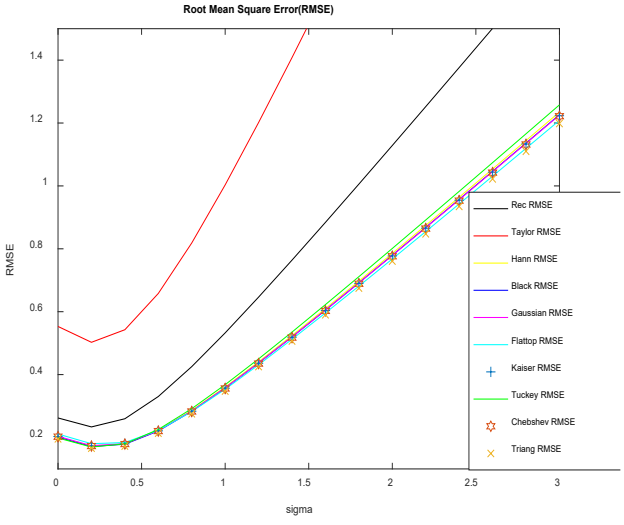
j) SAR image corresponds to Triangular window

Figure 9: Final SAR images of the examined windows

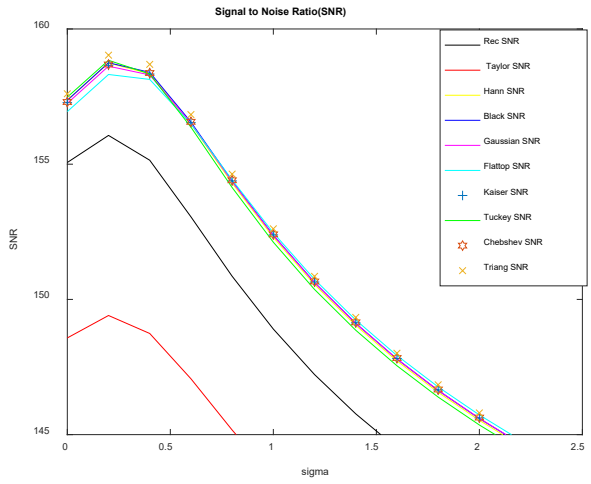
For the displayed images of Figure 9, MSE, RMSE, PSNR, and SNR are now calculated and the results are plotted in Figure 10.



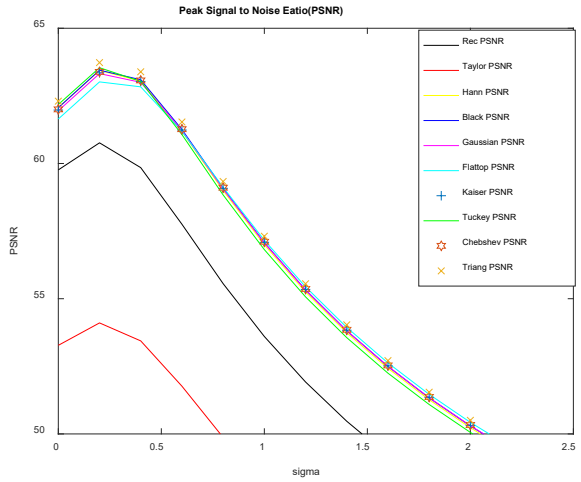
(a)MSE against σ



(b) RMSE variation versus σ



(c) SNR as a function of σ



(d) PSNR-σ behavior of different window functions

Figure 10:1 Noise sensitivity comparison of different window functions for earth image

From Figure 10, it is seen that Taylor window gives higher MSE and RMSE values, whilst its SNR and PSNR values are lowest. Owing to these worst values, Taylor window possesses the highest sensitivity to channel noise. Other windows perform better than the rectangular window and have less sensitivity to channel noise. Moreover, from Tables 2 & 3, it is recognized that PSLR of the SAR image can be improved by pulse shaping at the price of reducing the radiometric and spatial resolutions. Thus, the strategy of windowing, in general, enhances the image PSLR and consequently, it improves the system detectability in the range direction. On the other hand, the Tukey pulse shaping has the least enhancement factor due to its similarity to the rectangular window, however, its resolution is reduced by a factor of one and half relative to that of the rectangular window. Although Flattop image enhances the PSLR, its spatial resolution is lowered by 25% and its radiometric resolution is reduced by 2.7dB. So, the final image using Flattop pulse shaping has been appeared blurring and unclear as Figure 9 demonstrates. While the PSLR of Triangular window has the highest factor of improvement, which is 10.66 dB, its radiometric resolution is reduced by 1dB only and its spatial resolution is lowered to 54.5%. Next to Triangular window, the Gaussian pulse shaping comes. It has an enhancement factor of 6.72 dB in its PSLR image. However, its radiometric resolution is reduced by 1.6dB and its spatial resolution is lowered to 42.8%. Additionally, the Gaussian pulse shaping has the highest mean value which represents the high level of the reflected signal. Hann image reserves the next location. This type of window functions can similarly improve the image, as other windows, but its resolution is not reduced to a large value. Desbite Hann shows less PSLR improvement than Gaussian, its resolution performance is much better than Gaussian.

As indicated in [3], the Flattop window has the best performance, but this conclusion is from the SNR point of view only. However, studying the effect of this window on the resolution of the resulting image indicates that the final image has been appeared blurring and unclear. This means that the SNR only can't be used as a best way of selecting the suitable windowing category, but other factors should be taken into account, for any SAR system. On the other hand, the MIMO technique has been introduced in [9]. This technique improves the SAR image but, it requires high resources and its hardware is complex. Additionally, it has the demerits of increasing both the cost and power requirement. In [8], partial windowing has been suggested. This scenario of windowing has some cons as less energy-efficient windows.

In our study, PSLR is improved via windowing the transmitted pulse as well as using this windowing in the matched filtering process in the frequency domain. This has the benefit of easily handling the processing and simplicity in its practical applications.

6. Conclusions

This paper is concerned with studying the effect of windowing on SAR image recovery for the purpose of improving PSLR of such type of images. Different types of window functions are applied to two simulating images; point target and earth. From our displayed results, it is observed that Triangular, Gaussian and Hann give the highest performance for PSLR with different reductions in their resolution. The worst one is the Taylor due to its high sensitivity to noise. The remaining ones have an intermediate performance. Finally, it is concluded that the best pulse shaping is that selected in accordance with the design specifications of a SAR system.

References

[1] Oliver C. and Quegan S., "Understanding Synthetic Aperture Radar Images", Artech House, Boston, 1998.

[2] Y. K. Chan and V. C. Koo, "An Introduction to synthetic aperture radar (SAR)," *Progress In Electromagnetics Research B*, vol. 2, pp. 27–60, 2008

[3] Hyunguk Kim and Jinhwan Koh, "Analysis of Various Window Effect for SAR image Recovery," JIEIE, Vol. 52, No.12, December 2015, pp. 2108-2116.

[4] Li Jian and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging". IEEE Transactions on Signal Processing, 1996, 44 (6), pp.1469–1484.

[5] Bing Sun, Chuying Fang, Hailun Xu, and Anqi Gao, " A New Synthetic Aperture Radar (SAR) Imaging Method Combining Match Filter Imaging and Image Edge Enhancement", Sensors, Vol. 18, No. 12, 2018, pp. 1-16.

[6] Le-tian Zeng, Chun-hui Yang, Mao-sheng Huang, and Yue-long Zhao, "Verification of Imaging Algorithm for Signal Processing Software within

Synthetic Aperture Radar (SAR) System", Hindawi, Scientific Programming, Volume 2019, pp. 1-12.

[7] Birsen Yazıcı, Il-Young Son and H Cagri Yanik, "Doppler synthetic aperture radar interferometry: a novel SAR interferometry for height mapping using ultra-narrowband waveforms", Inverse Problems, Vol. 34, No. 5, May 2018.

[8] Byoung-GyunLim, Jea-ChoonWoo, Hee-Young Lee, and Young-Soo Kim, "A Modified Subpulsebpulse SAR Processing Procedure Based on the Range-Doppler Algorithm for Synthetic Wideband Waveforms", Sensors (Basel), 2008 Dec; 8(12): 8224–8236.

[9] Stéphane Méric and Jean-Yves Baudais, "Waveform Design for MIMO Radar and SAR Application", Topics in Radar Signal Processing, chapter 3

[10] Y. K. Chan and V. C. Koo, "An Introduction to synthetic aperture radar (SAR)", Progress In Electromagnetics Research B, Vol.2, 27–60, 2008

[11] R. Bamler, "Principles of synthetic aperture radar", Surveys in Geophysics, Vol. 21, pp. 147–157, 2000.

[12] Kazuo Ouchi, "Recent Trend and Advance of Synthetic Aperture Radar with Selected Topics", Remote Sens. 2013, Vol. 5, pp. 716-807

[13] Laurent Demanet, Matthew Ferrara, Nicholas Maxwell, Jack Poulson, and Lexing Ying, "A Butterfly Algorithm for Synthetic Aperture Radar Imaging", SIAM Journal on Imaging Sciences, 2012, Vol. 5, No. 1 : pp. 203-243

[14] A. Leukhina, V. Bezrodnyia, N. Parsaev, "Signals for Synthetic Aperture Radar", Procedia Engineering 201 (2017), pp. 398–407.

[15] Vishal M. Patel, Glenn R. Easley, Dennis M. Healy, and Rama Chellappa, "Compressed Synthetic Aperture Radar", IEEE Journal of Selected Topics in Signal Processing, Vol. 4, No. 2, APRIL 2010, pp. 244-254.

[16] Jun Shi, Yang Liu, Wei Liu and Xiaoling Zhang, "High-resolution synthetic aperture radar based on the IEEE 802.11 protocol", Electronics Letters, 2015, 51 (22). pp. 1815-1817.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 58-70 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

The Financial Services Authority of Indonesia E-Reporting System Development Based on Metadata XBRL Taxonomy

Rachmad Farizd[*,1], Suharjito[2]

[1]*Computer Science Department, Binus Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480,*

[2]*Computer Science Department, Binus Online Learning, Bina Nusantara University, Jakarta, Indonesia 11480*

ARTICLE INFO

Article history:
Received: 19 October, 2019
Accepted: 08 January, 2020
Online: 15 January, 2020

Keywords:
XBRL Taxonomy
Accounting Information System
Otoritas Jasa Keuangan (OJK)

ABSTRACT

The XBRL taxonomy has been proven to be able to restore various forms of redundancies and ambiguities in making financial reports from financial service institutions. However, the nature of XBRL which is oriented towards documents and based on XML makes the XBRL taxonomy difficult to use because the information cannot be directly read by ordinary users. The same obstacle was felt by the Financial Services Authority of Indonesia (Indonesian: Otoritas Jasa Keuangan or OJK) as financial regulator in utilizing the XBRL taxonomy, therefore a web-based application system called E-Reporting was developed to input financial statements conducted from financial service institutions to the regulator. The core of the application system is the XBRL taxonomy reading and mapping function, utilizing the Java library for reading XML and the Spring Java framework as a system backend, the Zkoss framework as the system frontend, and the database for storing data and specifically developed for the needs of the OJK. The purpose of this paper is to show other ways of implementing XBRL taxonomy, one of which is to present XBRL into a dynamic user interface according to the dynamic and flexible nature of XBRL. The results of the system evaluation indicate that the system is quite good at presenting the taxonomy into a user interface that suits the needs and can ease the burden of data validation, as well as providing convenience in forming the instance document, which is needed as the final result of the reporting process carried out by financial service institutions. Therefore, this system has been proven to provide added value in the use of XBRL taxonomy in financial reporting activities.

1. Introduction

The Financial Services Authority of Indonesia (Indonesian: Otoritas Jasa Keuangan or OJK) has the main duties governed by the law, namely in the field of supervision, monitoring, and decision-making. OJK establishes a one-stop integrated financial reporting policy as an accountability report for the financial services sector (SJK) so that a single point of contact can be realized for financial reporting from all financial services institutions (LJK). This reporting is based on metadata using the eXtensible Business Reporting Language (XBRL) taxonomy methodology so that the definition and reporting format for data uniformity is achieved, see [1]. The goal is to be able to reduce all forms of ambiguity and redundancy and facilitate the exchange and analysis of data that can have an impact on the quality of data that is useful in the decision-making process, see [2]. Although it is possible that the tag in the XBRL taxonomy is given incorrectly in financial terms, and allows the use of inconsistent standards, see [3].

OJK also has special attention to the weaknesses of document-oriented and XML-based XBRL. Making it difficult to exploit, process, adjust, search and retrieve data from a large set of XBRL files, even for tools that provide XML-based functionality, see [4] and [5]. Also, the process of manually retrieving and tagging data will spend hours and increase costs related to XBRL utilization, see [6, 7]. To overcome this weakness and facilitate the implementation of metadata-based reporting, a system was built as a pilot project in the reporting of the Financing Company (Indonesian: Perusahaan Pembiayaan or PP) called the SIPP client tool. This system is a substitute for taxonomy so that the process

[*]Corresponding Author Rachmad Farizd, Computer Science Department, Binus Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, 11530, Indonesia, Email: rachmad.farizd@gmail.co.id

of data retrieval and tagging can be done automatically then create XBRL instance documents that will be sent by LJK to OJK, so the taxonomy itself is only used as a reference as the formation of XBRL instances. This system has the following weaknesses and disadvantages:

- The system cannot follow the extensible nature of XBRL's taxonomy, which can be expanded and adapted to various variations and report requirements, see [8]. When there is a new version or an update from an existing version of the XBRL taxonomy owned by OJK, the system must be rebuilt and then distributed to LJK because the system is a standalone application.
- The input data validation process in the OJK client application is now in the form of a separate small application so that the input data must be re-validated on the client application in the LJK and the server in the OJK, so it is ineffective and inefficient.
- Large instance documents cannot be accommodated using a client application so that it can hamper the reporting process because it must be sent directly to OJK.
- Having the risk of financial information from the reporter being stolen because documents are sent through the internet network.

The implementation of the XBRL taxonomy in the SIPP client tool system is not optimal with these weaknesses and makes the vision and mission of the OJK unreachable in terms of reporting. The purpose of this research is to obtain and develop a replacement system to overcome the various weaknesses of the XBRL taxonomy mentioned earlier so that this metadata-based financial reporting can be applied throughout the FSS in the future. It is hoped that this new system can add value in the use of XBRL in terms of flexibility, efficiency, and security in the use of XBRL taxonomy in financial reporting from LJK to OJK.

2. Literature Review

All accounting information must be able to be issued clearly and effectively in digital form so that it can leave the traditional processes that are heavily exposed to human intervention. This is one of the obstacles and shortcomings of the use of XBRL for ordinary people because of the XBRL nature which is oriented towards documents and is based on XML, see [4]. The low level of understanding and knowledge of XBRL among accountants and other relevant stakeholders is also a significant problem as in [9]. The fact is that some respondents found refusing to use this new technology and the lack of application technology related software development to accommodate the use of XBRL, besides that the government also had to produce more training courses for entrepreneurs, see [10, 11].

There are several studies conducted to overcome the weaknesses in XBRL, thereby maximizing the implementation and adoption of XBRL. One of them is research by translating the XBRL format into Ontology Web Language (OWL) with the semantic web, the aim is to eliminate the possibility of redundancy and improve consistency so that conclusions can be drawn quickly and useful for business decisions of companies or organizations, see [5, 12]. Charles Hoffman also considers the semantic web to have many logical possibilities for the development of information technology so that it can produce more value, see [13]. The results are considered better in providing semantic data for XBRL, compared to previous research conducted by Garcia and Gil which presents a mapping based on structural transformation from XML to RDF (Resource Description Framework) and OWL, see [14]. However, OWL is not a language that can be directly used by business and financial people and not even for IT staff, besides OWL is designed only for special purposes so that the support of tools and expertise is minimal, see [15]. Table 1 is an example of the XBRL element transformation with the monetaryItemType data type to OWL.

Table 1 : Transformation of XBRL into OWL

Language	Transformation
XBRL	<element id="currentLoans" name ="currentLoans" type = xbrli:monetaryItemType xbrli:balance ="credit" substitutionGroup = "xbrli:item"> </ element >
OWL	Declaration (Class (ex:currentLoans)) SubClassOf (ex:currentAssets xbrlo:monetaryItemType) SubClassOf (ex:currentAssets xbrlo:credit)

Other research related to the development of portable open-source software, named Arelle to facilitate the reading of XBRL instances. The initial goal is to meet the need for devices that can parse and validate XBRL instance documents and support document versions. Arelle is an independent platform, developed entirely from scratch in Python. Using the MVC (Model-View-Controller) architecture, where the Model as a representation of XBRL objects, the Controller as a representation of interaction from users and program control from outside, and the View representation of the API interactions that were defined earlier with the Model to display the interface, see [16].

There is patent related to the development of systems and methods for processing, presenting, and mapping XBRL data to a set of flat tables, where each table represents a single hypercube projection. The aim is to maintain the very flexible nature of XBRL, and almost every entity that uses XBRL for reporting can extend the standard taxonomy to adjust its use to suit the needs of a particular entity. Because, this flexibility and adjustments can make it difficult to see data encapsulated in XBRL instance documents. The concept is to map XBRL data so that it can be mapped automatically back and forth between XBRL instances, by generating a set of flat tables automatically, where each table represents a single hypercube projection, see [17].

Other research related to the discussion and evaluation of the development of "XBRL-Passport" (Pilot Athens stock exchange's Statements Source Portal). This prototype was built as a smart

solution for clients, because of the long process and various formats of collecting and analyzing financial data from various sources, thus making companies have to spend more consistency costs, see [18]. The prototype architecture consists of 3 simple layers, namely: the database layer with MySQL, the logic layer with PHP, and the application layer using the web. Because XBRL data is embedded in various types of files such as XBRL, HTML, Excel, XML, or text, and the limitations of web applications for parsing the XBRL format, makes the information displayed in HTML format uses XBRL schemes rather than XBRL based. The initial targets of developing this prototype are:

- Provision of financial data storage locations from companies listed on the ASE (Athens Stock Exchange), by first being converted to XBRL.
- Make data relevant and ready to be analyzed.
- For companies with the same sector, comparisons can be made.

XBRL will not change the current operational standards and financial conditions or also significantly improve the quality of the company's financial statements. Companies may not get many benefits from implementing XBRL, even companies can spend more money to implement XBRL. But with XBRL it is possible to exchange financial information from a variety of different software platforms, see [7]. The requirements of a successful XBRL application are: (1) the specifications and taxonomies must be suitable for many companies, (2) application programs are similar to software that is currently widely used by accounting within companies, (3) can transfer information to specific forms were required by users. In [7], there are several reasons for implementing XBRL that can make companies spend more money:

- During the process of collecting, processing, and transferring, and investigating information requires a lot of money, see [19].
- Incomplete and misleading information releases can result in latency litigation costs (AICPA, 1976).
- By releasing information, the company can lose its competitive advantage, see [20].
- Companies can spend large funds to comply with regulations issued by the relevant government, and to ensure that information can be released or not, the company can re-issue costs, see [20].

3. Method

This study adopted the design science research (DSR) schema to describe the research design. The DSR framework can contribute to new knowledge when new solutions are proposed for known problems as seen in previous researches [21, 22]. The contribution of this study is the method of reading from the related XBRL taxonomy, then displaying it in another form of GUI that is flexible following the nature of XBRL. This research presents the development of a prototype system in the following steps: (1) gathering information related to the application of XBRL on an existing system; (2) uses information related to application to develop a process for reading taxonomies systematically; (3) changing the results of the XBRL taxonomy reading into a dataset that is ready to be transformed into an input form; (4) creating a dataset of links between data and tags from the related taxonomic schemes; and (5) output in the form of instance documents according to rules and needs.

System development will be carried out with the Agile methodology, which is a development methodology based on iterative development where requirements and solutions develop through collaboration between teams as in [23], to speed up the process and intensely involve stakeholders. The initial development was to create a function to reading and mapping the XBRL dictionary as the core system, and then save them into a function in the Java class. Facilitate the process of reading taxonomies and presenting taxonomy in the form of a GUI without having to make a special transformation into another form. Furthermore, the collection of information related to the system is done by observing the reporting process in OJK and conducting interviews with stakeholders in OJK. The data and information obtained will then be used as a reference and as input for the analysis process of the business needs and information expected by the OJK and the result will be a user and system specification document.

The design will use the help of UML (Unified Modeling Language) as a modeling tool commonly used to determine or describe a software system with objects, see [24]. The development of this system uses a web-based application architecture consisting of 3 simple layers: (1) the database layer, (2) the logic layer, (3) the presentation layer. Web application design will use the MVC (Model-View-Controller) pattern, making it easier for teamwork and workloads, because it separates the code into layers of presentation, logic, and data access. Separating the program code into 3 categories will make it easier to manage and read when there are parts of the code that need to be tested, improved, and modified, see [25].

After the system has been developed, it will then be tested using the User Acceptance Test (UAT), which is considered important and suitable for testing agile-developed software that allows visibility and improves communication and feedback to developers from end-user perspective or other stakeholder, see [26, 27]. This acceptance test can be sure whether these functions are acceptable to each other correctly, and the end of the test is carried out at the point between the completion of the development and the release system [28]. The UAT type used is Black Box Testing [29], by dividing it into 2 parts, first on the authorization side and the other on functional.

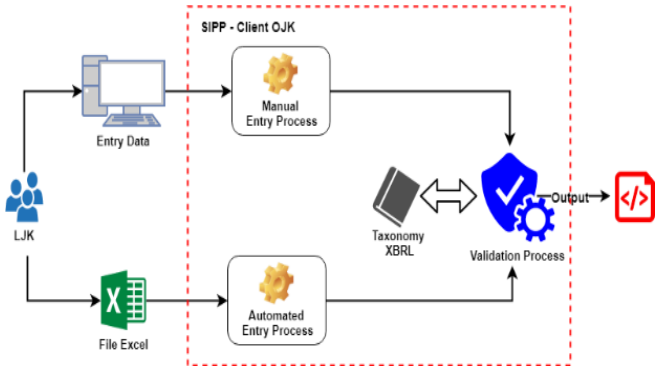


Figure 1: SIPP Client Current Process

4. Analysis and Results

4.1. OJK E-Reporting System

Figure 1 shows the implementation of the XBRL taxonomy in the SIPP client tool system which is not optimal as mentioned in the introduction. After the SIPP client system generates instance documents, LJK needs to send or upload the instance documents to OJK's Integrated Reporting System (SILARAS) for later validation. If the report passes the validation process, the document will be saved to the OJK file server, and if it is invalid then it will be rejected. The process of SILARAS can be seen in Figure 2.

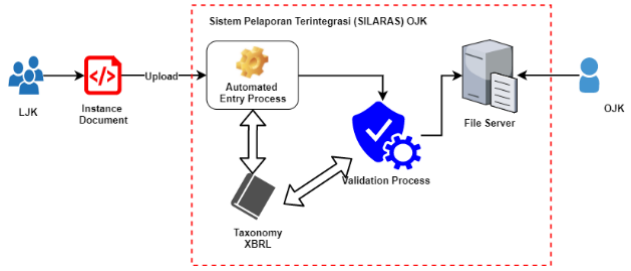


Figure 2: SILARAS Current Process

OJK's E-Reporting System is a web-based application system that has the main function is to read and map OJK's XBRL taxonomy for financial reporting conducted by LJK to OJK. The results of this taxonomy reading and mapping will be converted into GUI input forms that will be easier for reporters to use, even for users who are not familiar with XML and XBRL because they will no longer deal with XBRL taxonomy or XML formats. The system makes the validation process more efficient from the client side because the form that is displayed already has validation related to the input limits of each item that is converted into a component in the GUI form. Also, automation of creating report documents (instances) as output following general XBRL rules.

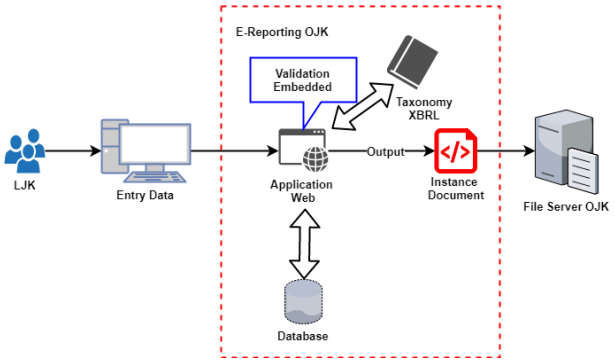


Figure 3: OJK E-Reporting system architecture

The focus of the discussion is how the system can read and map the taxonomy then present it into an input form for reporting and create an instance document as the result of the system. Also, the role of the validation process will be divided between LJK (client) and OJK (server) so that the validation process does not become redundant and more efficient in time and resources. Another plus is that all data inputted and instance documents have been stored on the OJK server so that the LJK does not need to upload instance documents which are usually quite large and take a long time and can increase the security aspects of financial information of LJK as the reporter as shown in Figure 3.

With instance documents already contained in the OJK file server, LJK does not need to send them to the SILARAS application. OJK only needs to select the instance documents that need validation, and the validation is only related to formulas as shown in Figure 4.

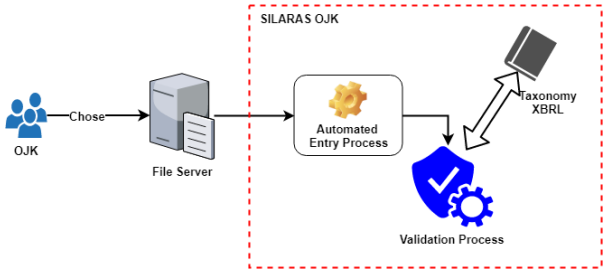


Figure 4: Architectural design application SILARAS OJK

The system architecture will consist of 3 simple layers: the database layer will use the MSSQL Server 2012 database, the logic layer uses the Java programming language with the Spring framework, and the presentation or user interface layer uses the Zkoss framework. Java was chosen because it has a complete library that can be used for reading XBRL taxonomies based on XML, although not specifically for XBRL taxonomies, so there is no need to convert to OWL or RDF as in existing research on semantic web [5, 12, 13]. Zkoss was chosen because it has been proven from several application system developments conducted by the author, it can present and support dynamic user interfaces according to system requirements. Also, Java and Zkoss have been used by OJK in developing application systems as well. The XBRL taxonomy architecture used is existing and standard used by the OJK with a folder hierarchy and the naming of the taxonomies determined by the OJK according to their needs.

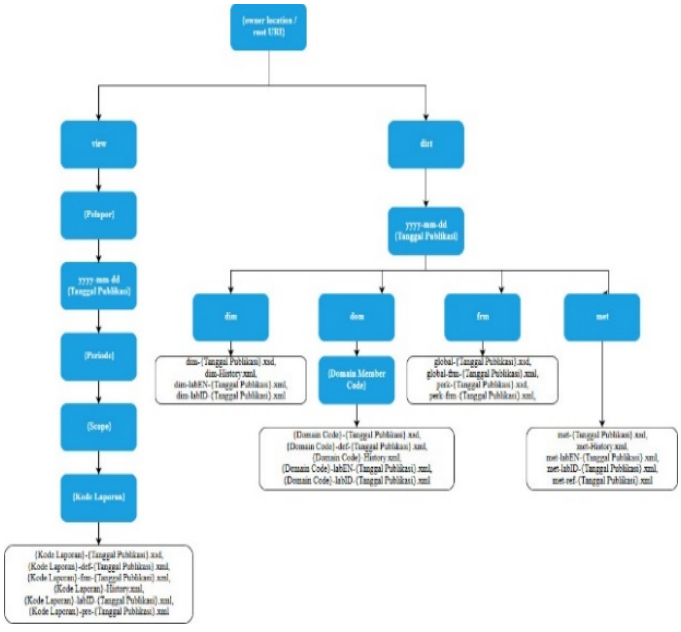


Figure 5: The folder and file structure of the OJK XBRL taxonomy

4.2. Root XBRL Taxonomy of OJK

The root (aloc) is the prefix of the address or location of the XBRL file, even though the file is actually in a local folder on the server computer commonly referred to as a rewrite URL. Aloc contained within the taxonomy framework used by OJK in each file are: http://xbrl.ojk.go.id/taxonomy/.

4.3. OJK Taxonomy Folder and File Structure

OJK has established a folder structure and component framework to facilitate the design and grouping of taxonomies used in financial reporting by LJK to OJK. Under the root folder, there are 2 main folders, namely 'dict' and 'view' as shown in Figure 5.

- The 'dict' folder, also known as a dictionary, contains all the entities, dimensions/axes, and domains needed for all groups of information. The modeling uses Data Point Model (s) (DPM) published by The European Banking Authority (EBA). In the dictionary there are metrics (met), dimensions (dim), and domain (dom) which are grouped into the release date folder as shown in Figure 5.
- The 'view' folder contains the main schema file that was first read, and there are several folder hierarchies intended to help classify the report. In the OJK, the classification or report categories are divided into reporters, publication date, period, and scope for each report as shown in Figure 6, with reading patterns:

{root URL}/view/{reporter}/{publication date}/{period}/{scope}/{report code}/{file name}

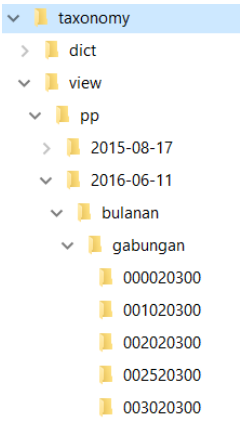


Figure 6: The structure of the view folder on OJK's taxonomy

4.4. Core Process in OJK's E-Reporting System

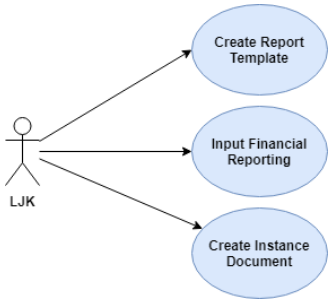


Figure 7: Use case in OJK's E-Reporting system

In the E-Reporting system there are 3 main entities that users (LJK) do to the system related to the XBRL taxonomy as shown in Figure 7.

- Creating a report template.
- Fill out the report.
- Creating instance documents.

Furthermore, all these entities will become the core processes in the OJK E-Reporting system related to the XBRL taxonomy as shown in Figure 8.

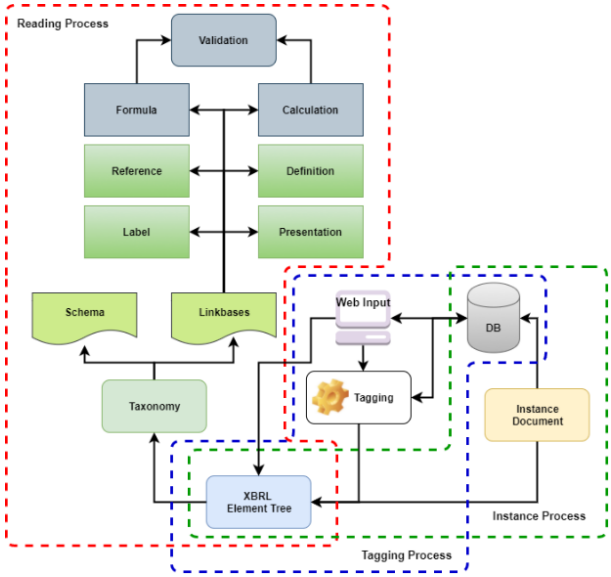


Figure 8: The core process in the OJK E-Reporting system

4.5. Taxonomy Reading and Presentation Process

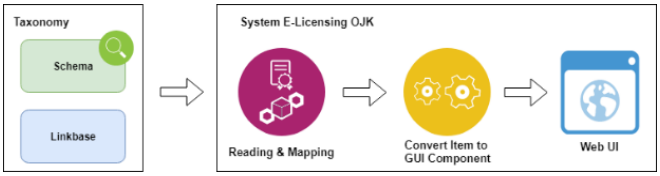


Figure 9: Taxonomy reading and transformation process in the OJK E-Reporting system

In Figure 9, the components of the taxonomy are divided into two, namely the schema and linkbase. Schema is part of the taxonomy that defines the items that will be used in reporting, along with the type, and data structure of information. Whereas linkbase is a logical expression of the relationships and relationships between items and the standards that govern it and information specifically related to the scheme. In the reading process, the schema which is the main structure of the taxonomy is stored in a file with the xsd extension. The schema in the view folder is the first file to read, with the pattern as explained before. The naming of this schema file has a naming pattern:

{Report Code}-{Date Version}.xsd.

Schema file contains metadata elements related to the report, the element is part of the taxonomy that contains items and other data related to the items that are linked by linkbase as shown in Figure 10.

element	ElementDeclImpl (id=281)
abst	true
id	"F000020300_abs1" (id=293)
m_annHolder	AnnotationHolderImpl (id=294)
m_elem	1
m_elementNode	5366
m_isAbstSpecified	true
m_isMaxOccursSpecified	false
m_isMinOccursSpecified	false
m_isNillableSpecified	true
m_item	0
m_modified	false
m_name	"abs1" (id=304)
m_numeric	0
m_registered	true
m_relLinkList	LinkListImpl (id=308)
m_schema	XMLSchemaImpl (id=310)
m_simpleType	null
maxOccurs	1
minOccurs	1
nillable	true
periodType	"duration" (id=324)
subGroupLocalName	"item" (id=326)
subGroupNamespaceURI	"http://www.xbrl.org/2003/instance" (id=328)
typeLocalName	"monetaryItemType" (id=330)
typeNamespaceURI	"http://www.xbrl.org/2003/instance" (id=328)

Figure 10: The results of reading the taxonomy elements in the OJK E-Reporting system

In the process of reading and mapping taxonomy XBRL, the system will read directly the elements of the taxonomy selected, but not converted into OWL or RDF as previous research. In the system already available Java class to accommodate the results of taxonomic readings, the reading and mapping process can be faster than having to convert it first into other forms. In general, the reading of the taxonomy carried out by the system for the relationship between documents can be seen in Figure 11.

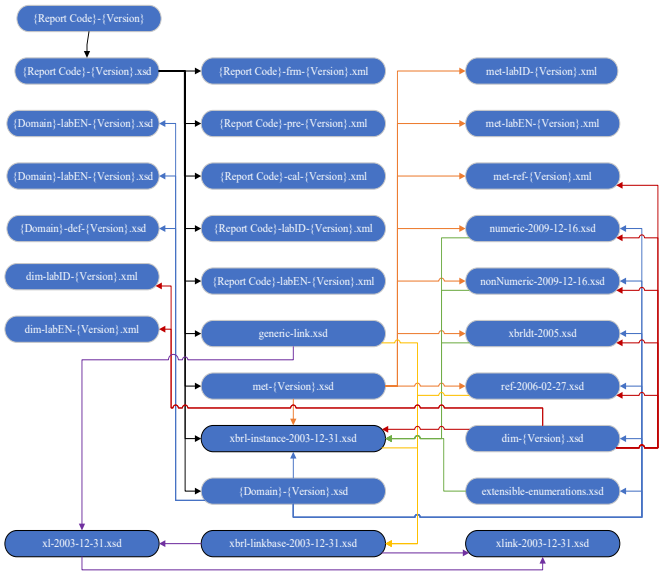


Figure 11: Relationships between files in taxonomy in general

To support the process of reading taxonomies and storing report template data, a data model was made to support the implementation of taxonomies in the OJK as shown in Figure 12. Information stored in XBRL documents cannot be accessed easily for safe and scalable queries and analyzes. So sometimes companies make two copies of data that have the same information, and one of them is stored in a relational database other than the XBRL document itself [30].

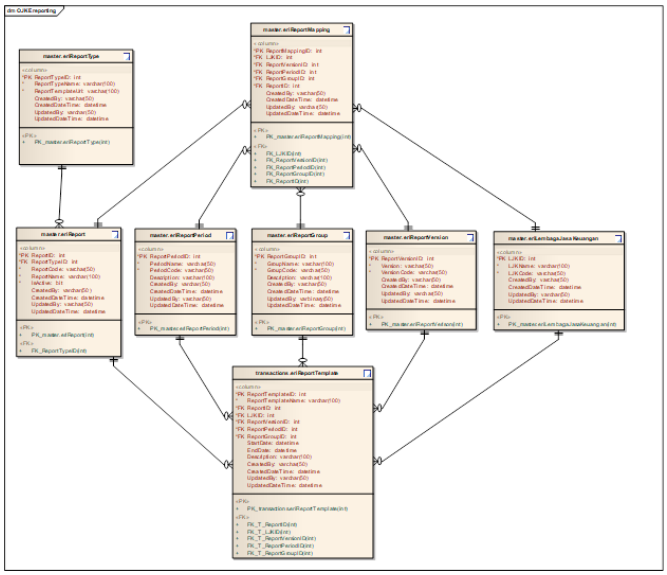


Figure 12: Data model for OJK E-Reporting system

The user must first create a report template based on the taxonomy associated with the report. The reports in OJK are grouped into several criteria and become a folder structure within the taxonomic storage folder as explained in the folder structure points above. This is intended to facilitate the user in viewing the report history or create adjustments to the data if needed. The draft input form of the report template as shown in Figure 13, will then be displayed entirely in the report template list module in the system as shown in Figure 12.

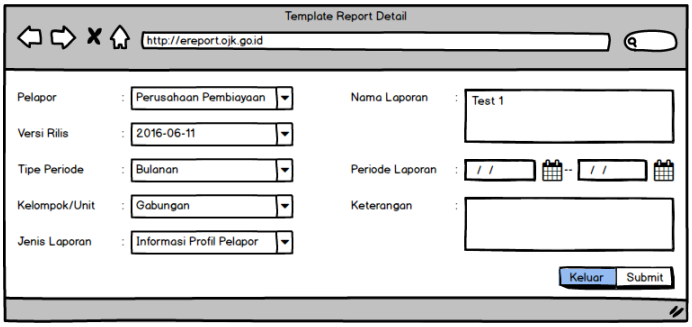


Figure 13: Design input forms for report templates in the OJK E-Reporting system

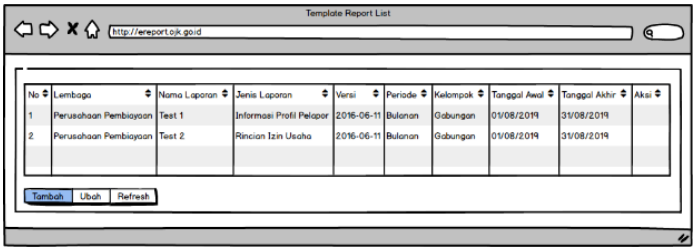


Figure 14: Design list of report templates in the OJK E-Reporting system

In the first part of the schema reading, you will find a namespace with a prefix representing it. The namespace in XBRL is no different from what is understood in XML, namely the naming mechanism that uniquely groups elements and attributes. In the system, the namespace will always be carried as a reference

in the creation of an output report document (instance). In Table 2, you can see a list of common namespaces used in the XBRL taxonomy at OJK.

Table 2: The namespace contained in OJK's XBRL taxonomy

Prefix	Namespace
Xsd	http://www.w3.org/2001/XMLSchema
Gen	http://xbrl.org/2008/generic
Ref	http://www.xbrl.org/2006/ref
Xbrldt	http://xbrl.org/2005/xbrldt
Nomnum	http://www.xbrl.org/dtr/type/non-numeric
Xl	http://www.xbrl.org/2003/XLink
Link	http://www.xbrl.org/2003/linkbase
Num	http://www.xbrl.org/dtr/type/numeric
Xlink	http://www.w3.org/1999/xlink
Enum	http://xbrl.org/PR/2014-03-26/extensible-enumerations
Xbrli	http://www.xbrl.org/2003/instance
Xsi	http://www.w3.org/2001/XMLSchema-instance
iso4217	http://www.xbrl.org/2003/iso4217
variable	http://xbrl.org/2008/variable
F{Kode_Laporan}	http://xbrl.ojk.go.id/taxonomy/F{Kode_Laporan}

Besides that, in the scheme there are also public elements, which are other concepts of the model represented by codes that are within a certain scope and can provide additional information from key elements such as readable label information, the definition of elements, and references other legal in languages that might be different. Information regarding this resource is generally marked as XLink and represented as an XBRL item, as shown in Figure 15.

```
<link:linkbaseRef xlink:type="simple"
  xlink:href="000020300-pre-2016-06-11.xml"
  xlink:role="http://www.xbrl.org/2003/role
/presentationLinkbaseRef"
  xlink:arcrole="http://www.w3.org/1999/xli
nk/properties/linkbase"/>
```

Figure 15: Examples of public elements in the OJK XBRL taxonomy

All resources will have their respective roles following the reference linkbase, also can have their own rules. Besides, the different roles make it possible to apply the Codification itself to present the resource, for example for label resources, the code is 'lab', while for the resource definition is 'def', and for the presentation is 'pre'. More complete coding of the linkbase in the OJK taxonomy can be seen in Table 3.

Table 3: Naming on linkbase in the OJK XBRL taxonomy

Linkbase	Name
Label	[report code]-lab[lang]-[version date].xml
Presentation	[report code]-pre-[version date].xml
Reference	[report code]-ref-[version date].xml
Calculation	[report code]-cal-[version date].xml
Formula	[report code]-frm-[version date].xml

Reading the schema will produce elements along with additional information contained in the schema. Then the author tries to simplify the reading of elements, and compile them into a hierarchy with the key in the form of the unique identity of the elements so that it is easy to read and transformed into an input form in the system, henceforth the authors refer to a set of data from these elements into node elements. Examples of node element hierarchies for reading taxonomy schema can be seen in Figure 16.

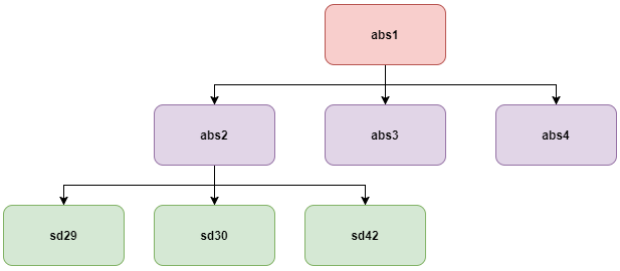


Figure 16: Examples of tree elements from taxonomic reading

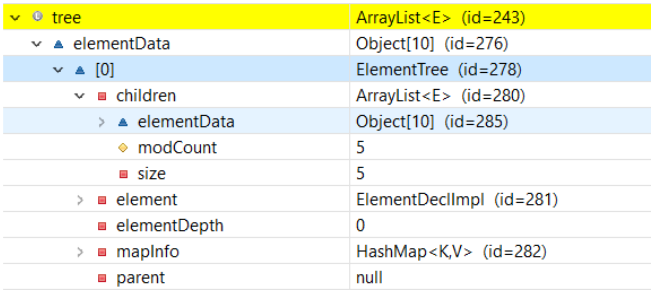


Figure 17: Example of element node results on the OJK E-Reporting system

As shown in Figure 17, at each node the element will have the same attribute property but with different values according to the role. In the reading of the scheme will be divided into 2 (two) types, namely items and tuples. Items do not have nodes underneath, whereas tuples can have one or more items underneath. For example, abs1 is an absolute element that indicates that this node is a tuple and has other derived elements in it. Properties inside the node can be seen in Table 4.

Table 4: Attributes on element nodes

Attribute	Description
Children	List of nodes that are 1 level below
Element	The XBRL metadata is specific to the node
MapInfo	Summary of information from element that are often read and used in the presentation process
Parent	Summary of information from nodes that are 1 level above

In this presentation process, only the 'Children' and 'MapInfo' attributes are read because these 2 attributes are sufficient to have the information needed to transform the taxonomy schema that has been read into an input user interface in the system. If an element has a derivative or is an absolute node, then it can be certain that it is an information group or tuple. The process will sort elements that do not have children and are not tuples to be transformed into input components in the form to be displayed.

The type of input will be different according to the type of the element, and the different and need special treatment is the type of enumeration element, because this type of element will look to a related domain. The types of elements and components used can be seen in Table 5, while the conversion process is shown in Figure 18.

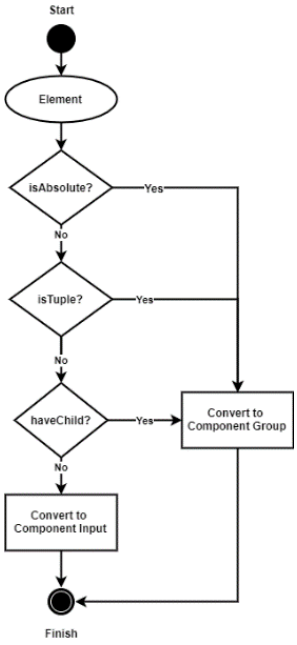


Figure 18: Flow conversion of elements into GUI components on the OJK E-Reporting system

Table 5: Conversion of elements into GUI components

Type Data	Component Input Type
monetary	Integerbox
decimal	Integerbox
integer	Integerbox
date	Datebox
time	Timebox
string	Textbox
enumeration	Combobox

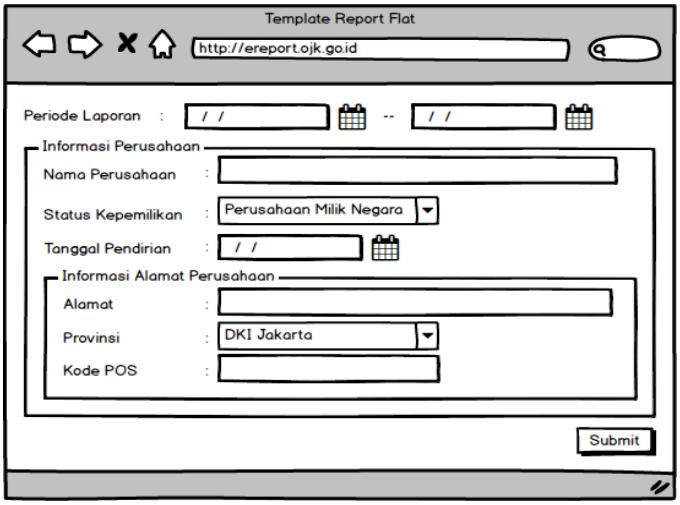


Figure 19: Flat type input form on OJK E-Reporting

For the user interface on the input form, the appearance is different for each type of form (flat, tuple, and nested tuple). Flat type will display a form like most simple input forms where there are labels and component inputs that accommodate data input such as text, date, number, and others. Example of the results of converting taxonomy items with flat type into a user interface component that will be displayed, it can be seen as shown in Figure 19.

For other types of input forms, they will be presented as data tables as shown in Figure 20. The form details will appear as popups if data is added or changed, and the shape is the same as flat types.

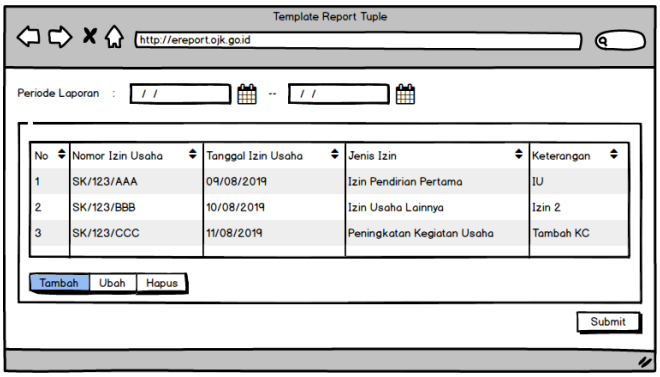


Figure 20: Tuple and nested tuple data input form types on the OJK E-Reporting system

Inside 'MapInfo' there is some information such as identity, label, and constraints such as maximum length, format, and what can be input into the component. All this information is in the XBRL results from the reading. With this limitation already embedded in the input form component so that the validation process becomes lighter, for example in the input component with a numeric type, the characters that can be inputted only in the form of digits with a digit length are following the taxonomy reading results. So that this presentation process has fulfilled the dynamic requirements of the user interface and validation of input components.

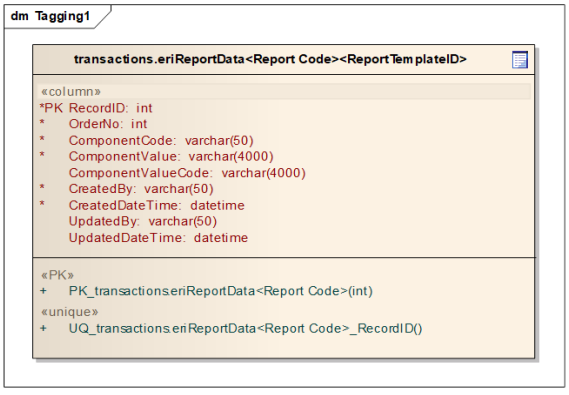


Figure 21: Data model for the results of data tagging on the OJK E-Reporting system

4.6. Data tagging process

The tagging process only stores the values of the input components on the form with the element id. The difficulty is to

provide a data model that can accommodate all types of forms and support any changes that exist, both the addition and subtraction of elements from the taxonomy. For this reason, a data model design has been made that can accommodate existing needs, from the flat, tuple, and nested tuple form types having the data model scheme as shown in Figure 21.

The database table is contained in the transaction table data schema and will be created automatically for each report template created when the taxonomy for the template is read, making it easier and faster to query data tagging because data queries go directly to specific tables for data- data related to the financial statement template created. The table name will have a pattern:
transactions.eriReportData<Report_Code><Report_Tmpl_ID>
Although the taxonomy will change in the future, this data model will not be affected because the system will only synchronize the tagging data only because the main orientation of the data tagging is still based on the results of the XBRL taxonomy reading.

4.7. Creating instance documents

An instance document is a file containing data reported based on the XBRL taxonomy consisting of a schema and linkbase file, which has been created by OJK. In this OJK E-Reporting system, instance document files will automatically be formed after the taxonomy presentation process and the data tagging process have been carried out. The instance creation process uses the data stream method so that it does not require large memory. The process of forming documents has been based on XBRL instance document provisions in general and stored on the file server repository. The instance document file in OJK has the form shown in Figure 22.

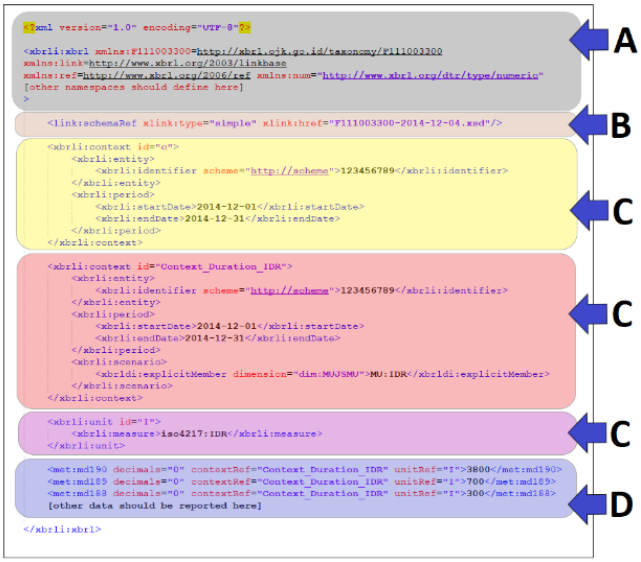


Figure 22: Example contents of instance document content generated by the OJK E-Reporting system

The following is an explanation of each part of the instance document:

- The header consists of XML Header and XBRL Header. In the XBRL header, the namespace and schema location is listed for the namespace used that is used in the document instance and is usually the same as the taxonomy (A).
- Schema location as described previously outside the instance document informs about the taxonomic reference used (B).
- The context consists of elements that contain information about the reporter and contain information about the reporting period. Also, scenarios can be added, namely specific conditions obtained from the intersection of values of the dimensions involved (C).
- Primary items are items that contain reported data that has been tagged with references that are needed. For primary items that are empty, the 'xsi:nil="true"' attribute is added (D).

5. Experimental Result

. The test only uses a representative of 1 report with flat type, not on all reports in PP. The focus of the experiment is how the system presents reports from the reading of taxonomy schema, until instance documents are formed. Experiments carried out from the user to log in, until the formation of report documents known as document instances. The login module in Figure 23, is the login standard that exists in all OJK systems. Login can use system validation by checking the master user table, and using LDAP validation. To add to the safety factor, a caption check is added, to ensure that no other system is attempting to enter the OJK E-Reporting system.

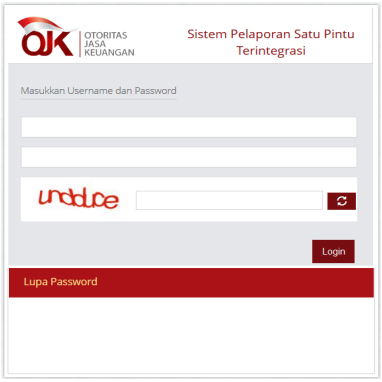


Figure 23: Login module on the OJK E-Reporting system

Figure 24 at record number 1. Input form to add report templates as shown in Figure 25, which has been adjusted to the needs of financial statements and taxonomy structure to the OJK. The input form for the list report template is also adjusted to the architecture of OJK's taxonomy, according to the design of the previous system.

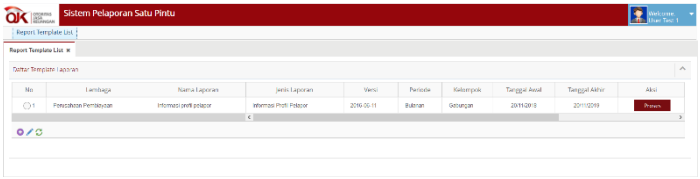


Figure 24: List of report templates module on the OJK E-Reporting system

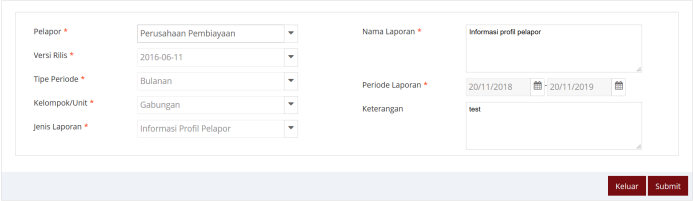


Figure 25: Report template input form module in the OJK E-Reporting system

Figure 26 shows the folder structure and the files associated with the selected report template. The folder presents the taxonomic grouping structure applied in OJK and becomes the parameter setting of the report template. The first schema file to read is a file with XSD extension, as explained at the taxonomy reading process point.

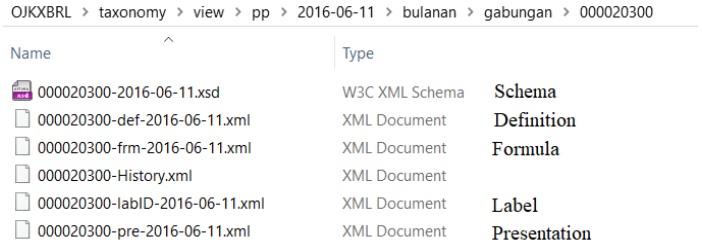


Figure 26: The folder structure and the files associated with the selected report template

In the schema file, the system will read all namespaces and linkbases associated with this scheme. Most of the main files needed are in this folder, except for the metadata associated with the dictionary (dict) as explained in the previous taxonomy reading point. In this experiment, we used samples for absolute elements (abs1 and abs2) as shown in Figure 27, to minimize the scope of the experiment. As seen in Figure 11, to see the child elements of these items, the system will read the presentation file form the folder that is read in Figure 26.

Figure 27: Example schema file that was read in this experiment

In the presentation file that is read, we can see that the abs2 item is the child of abs1 item. And abs2 has several derived items in it as seen in the presentation file in Figure 28. From here, the system can read the required data set and will read other related metadata from items that will be displayed such as labels, dictionaries related to boundaries, and the association of data with other dimensions. The system will process the taxonomy reading result of the metadata, then display it in the form of a user interface. That way XBRL users don't need to deal with taxonomy in XML anymore, but already in the GUI format as shown in Figure 29.

Figure 28: Example presentation file that was read in this experiment

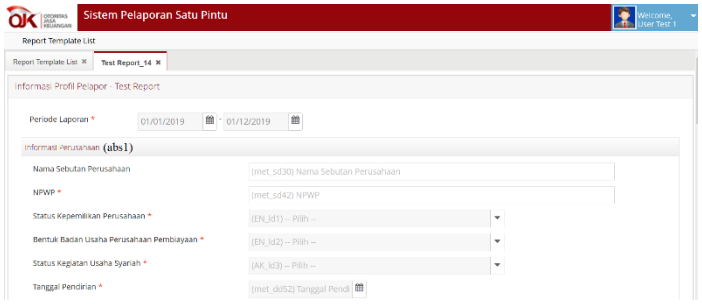


Figure 29: Report input module for flat type in OJK E-Reporting system

The experiment is continued by making changes to the taxonomy, for example by removing items or changing labels and input limits related to the selected report. In the presentation file in Figure 28, items with identity sd42 are subtracted. Then the label of item sd30 is changed, and the input given a restriction must be filled. Figure 30 shows an experiment related to change as mentioned before. Items with the identity sd42 are no longer found in the form. Whereas for items with sd30 there are changes to the label, and the input has a different data validation from before which is not allowed to be blank as shown in Figure 30. Validation of input components includes checking the blank data, allowed characters, and the length of characters inputted. Experiments show system boundaries related to this validation, the system cannot display data that has been covered if changes are made in the taxonomy related to this limit, so the data does not pass the validation. As a result, the system is well adapted to changes in both the data and taxonomy structure. If there are items that are omitted in the taxonomy, then the data that has been inputted cannot be maintained in the database.

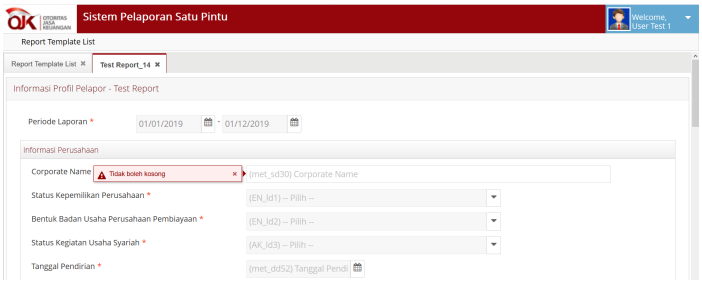


Figure 30: Report input module after taxonomy modification

After the user input data, the next task of the system is to tag data. The goal is that data can be added into a dataset that will be ready to be written into an instance document as shown in Figure 31, as result of the process. Experiments show the system can properly produce instance documents even though the taxonomy

has changed, both structurally and in the dictionary associated with the items contained in the taxonomic structure. Regarding the experiment mentioned earlier, when an element with identity sd42 is removed from the presentation an instance document will be formed as shown in Figure 32. It is important to understand, only changes in presentation and data types will affect document instances.

Figure 31: Document instance generated before taxonomy modification

Figure 32: Document instance generated after taxonomy modification

Table 6: UAT scenario on authorization

Date	*Title*	*Description*
	Login	Testing the login module
Case-Id	*Description*	
LOG-01	Log in by filling in the username column	
LOG-02	Log in by filling in the password field	
LOG-03	Log in by filling in the captcha column	

Table 7: Potential errors on the UAT authorization

Case-Id	*Scenario*	*Screen*	*Sub-Screen*	*Problem*
AU-01	Log in to the application	Login	01	Username, password and captcha are left blank
			02	Incorrect username, password and captcha

6. Testing and Evaluation

OJK E-Reporting System is tested and evaluated using the User Acceptance Test (UAT) at the Directorate of Statistics and Information IKNB (DSIN) and the OJK Financing Company Supervisor. Acceptance testing performs validation on the software product, and UAT is one of the classified of acceptance tests. The UAT will focus on the requirements of the system design based on the system use case scenario to consider the usability of the system and ensure that the process is carried out in the right way [27-29]. Testing goes on by dividing it into 2 parts, namely: authorization and functionality. For UAT authorization only focuses on access rights to report templates that are owned only. Test scenarios for authorization testing can be seen in Table 6, and error potential in Table 7.

After doing the UAT on the authorization side, no errors were found in user authentication, but there was an error in the list of report templates that were displayed not in accordance with their rights. Furthermore, in functional testing, test scenarios can be seen in Table 8 and potential errors in Table 9.

Table 8: Skenario UAT on functionality

Date	*Title*	*Description*
	Form	Tests for testing taxonomy transformations into input forms for data reporting
Case-Id	*Description*	
FRM-01	Select a report template to display as an input form	
FRM-02	Entry data and then saved	

Table 9: Potential error UAT on functionality

Case-Id	*Scenario*	*Screen*	*Sub-Screen*	*Problem*
UI-01	Data reporting	Form	01	Displays report forms that have no taxonomy
			02	Entry data with incorrect data

After conducting the UAT on the fictional presentation of XBRL taxonomy and input data, no errors were found. It needs to be re-validated if there are report templates for which there are no taxonomic guidelines, then they don't need to be displayed. The process of presenting taxonomies into input forms also takes a little time, but the forms already have input limitations so the process of making instance document reports becomes faster. Finally, an evaluation of the system readiness carried out by the stakeholders to ensure the system is ready can be seen in Table 10.

Table 10: OJK E-Reporting system readiness checklist

Items	Description	Values
Scope of development	The scope of development consists of re-engineering a system of existing systems that are live and used.	☒ Good ☐ Enough ☐ Not good
Unit test by developer	A unit test has been carried out by the developer.	☒ Good ☐ Enough ☐ Not good
User Acceptance Test	• User Acceptance Tests are carried out in the main functions of the system as a whole because the system was	☐ Good ☒ Enough ☐ Not good

	developed from scratch, with reference to the existing system using the Application Making Request Form. • Testing has been carried out both by the Directorate of Statistics and Information of the IKJB (DSIN) of the OJK and the OJK Financing Company Supervisor (May 19, 2019) with good results and all features running smoothly.	
Primary & secondary DC locations	Main server location: DC colocation at Sentul Bogor, Indonesia	☐ Good ☒ Enough ☐ Not good
Hardware specifications	• 3 Web Servers • 1 Database Server Server virtual with high availability	☒ Good ☐ Enough ☐ Not good

7. Conclusion

From the results of discussions that have been obtained in the development and testing of the OJK E-Reporting system, it can be concluded as follows:

- The system has been able to properly present the XBRL taxonomy flexibility adopted by XBRL, although it cannot yet be used for complex financial reporting of LJK.

- Separation of validation has a big impact on processing time because there is validation on the client-side that has been done when data is inputted. This has an impact on the processing of overall validation at OJK so that it becomes faster and more effective.

- Automated creation and storage of XBRL instance documents performed by the OJK E-Reporting system does not overload the client-side (LJK), because it is directly created and stored on the server-side (OJK).

OJK's E-Reporting System can cover weaknesses and deficiencies in the existing system (SIPP client tool), in terms of flexibility, validation efficiency, and convenience. The system still has great potential to be developed because it cannot yet support the reading of complex taxonomies with the implementation of interrelationships between existing reports. Validation placement based on formulas will no longer be needed if the input component in the taxonomy transformation form can be checked directly when the user fills into input data. Also, it is necessary to add varieties of input data types (e.g. xls file), so that users can provide data in other forms and do not need to manually input data.

References

[1] J. W. Kim, J.-H. Lim and W. G. No, "The effect of first wave mandatory XBRL reporting across the financial information environment.," *Journal of Information Systems,* vol. 26, no. 1, pp. 127-153, 2012.

[2] R. Chowdhuri, V. Y. Yoon, R. T. Richard and U. O. Etudo, "Ontology based integration of XBRL filings for financial decision making," *Decision Support Systems,* vol. 68, pp. 64-76, 2014.

[3] A. Bahri, "Pros and Cons of XBRL (A New Era In The Field of Commerce)," *International Journal of Business Management and Scientific,* vol. 30, June 2017.

[4] J. Bao, G. Rong, X. Li and L. Ding, "Representing Financial Reports on the Semantic Web," *Semantic Web Rules Lecture Notes in Computer Science,* pp. 144-152, 2010.

[5] R. García and R. Gil, "Publishing XBRL as linked open data," *In CEUR Workshop Proceedings,* vol. 538, 2009.

[6] M. Vasarhelyi, D. Chan and J. Krahel, "XBRL consequences to financial reporting, data analysis, decision support, and others," *Rutgers University,* 2010.

[7] H. Chen and W. Sun, "Study on Rational Application of eXtensible Business Reporting Language," *JSW,* vol. 6, no. 2, pp. 257-264, 2011.

[8] M. Willis, "Improving investor communications and analysis via standardization," *The International Journal of Digital Accounting Research,* vol. 7, pp. 153-165, 2007.

[9] S. A. A. Azhar and U. Subramanian, "Impact of XBRL in Emerging Countries," *Available at SSRN 3462951,* 2019.

[10] S. R. Abed, "The Perception of XBRL Technology in the Jordanian Context: An Exploratory Study," *Research Journal of Applied Sciences,* vol. 13, no. 1, pp. 1-4, 2019.

[11] C. E. L. Ceballos, "The International Language of Information Xbrl, in the Tuluá SMES," *International Journal of Economics and Management Systems,* vol. 3, 2018.

[12] C. Hoffman and M. M. Rodríguez, "Digitizing Financial Reports – Issues and Insights: A Viewpoint," *The International Journal of Digital Accounting Research,* vol. 13, 2013.

[13] C. Hoffman, "Financial Reporting Using XBRL, IFRS and US GAAP Edition," *UBMatrix,* 2006.

[14] R. García and R. Gil, "Triplificating and linking XBRL financial data," *Proceedings of the 6th International Conference on Semantic Systems,* September 2010.

[15] J. Cardoso, M. Hepp and M. D. Lytras, The semantic web: real-world applications from industry, vol. 6, Springer Science & Business Media, 2007.

[16] H. Fischer and D. Mueller, "Open Source & XBRL: the Arelle® Project," *In 2011 Kansas University XBRL Conference,* pp. 29-30, 2011.

[17] C. Binstock and B. Milnes, "XBRL flat table mapping system and method". Patent U.S. Patent No. 9,292,544, 22 Mar. 2016.

[18] E. Stergiaki, A. Vazakidis and A. Stavropoulos, "Development and Evaluation of a Prototype web XBRL-Enabled Financial Platform for the Generation and Presentation of Financial Statements according to IFRS," *International Journal of Accounting and Taxation,* vol. 3, no. 1, pp. 74-101, 2015.

[19] R. K. Elliott and P. D. Jacobson, "Costs and benefits of business information," *Accounting horizons,* vol. 8, no. 4, pp. 80-96, 1994.

[20] R. L. Watts and J. L. Zimmerman, "Positive accounting theory: a ten year perspective," *Accounting review,* pp. 131-156, 1990.

[21] S. Gregor and A. R. Hevner, "Positioning and presenting design science research for maximum impact," *MIS quarterly,* pp. 337-355, 2013.

[22] C. C. Chou, C. J. Chang and J. Peng, "Integrating XBRL data with textual information in Chinese: A semantic web approach," *International Journal of Accounting Information Systems,* vol. 6, pp. 32-46, 2016.

[23] R. S. Pressman, Software engineering: a practitioners approach, Boston: McGraw-Hill Education, 2015.

[24] L. D. Bentley, K. C. Dittman and J. L. Whitten, System analysis and design methods, Boston: McGraw Hill, 2000.

[25] L. Forte, "Building a Modern Web Application Using an MVC Framework," 2016.

[26] B. Haugset and G. K. Hanssen, "Automated acceptance testing: A literature review and an industrial case study," *In Agile, 2008. AGILE'08. Conference,* pp. 27-38, August 2008.

[27] P. Pandit and S. Tahiliani, "AgileUAT: A framework for user acceptance testing based on user stories and acceptance criteria," *International Journal of Computer Applications,* vol. 120, no. 10, 2015.

[28] J. A. Shim, H. J. Kwon, H. J. Jung and M. S. Hwang, "Design of acceptance test process with the application of agile development methodology," *International Journal of Control and Automation,* vol. 9, no. 2, pp. 343-352, 2016.

[29] M. E. Khan, "Different Software Testing Levels for Detecting Errors," *International Journal of Software Engineering (IJSE),* vol. 2, no. 4, pp. 70-80, 2011.

[30] I. Belev, "Alternatives for Storing and Validating XBRL Data," *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS),* vol. 60, no. 1, pp. 191-201, 2019.

ASTES Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 71-76 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Evaluation of the Quality Parameters of a 4G-LTE Communications Base Station, Installed in a Rural Area of Peru

Omar Chamorro-Atalaya[1,*], Eduardo Pizarro–Mayta[2], Dora Arce-Santillan[1]

[1]*Faculty of Engineering and Management, Universidad Nacional Tecnológica de Lima Sur, Perú*

[2]*Faculty of Electronic and Computer Engineering, Universidad Nacional Federico Villarreal, Perú*

ARTICLE INFO

Article history:
Received: 24 November, 2019
Accepted: 22 December, 2019
Online: 15 January, 2020

Keywords:
Quality parameters;
Communications Base Station;
4G-LTE technology; Rural zone;
International
Telecommunications Union.

ABSTRACT

This article has to specify the quality parameters of a Communications Base Station with 4G-LTE technology, installed in a rural area of Peru. The problem of this investigation is due to the fact that many areas of the country do not have a communications network capable of covering the high demand of the inhabitants; this lack ultimately generates a gap in the capacity of information transfer and communications; since, since there is no connectivity between the inhabitants, it is very difficult to have access to information, and generate economic opportunities that allow the growth of this area, affected rural. To determine the quality of the implementation of a 4G-LTE network, it was sought to specify the characteristics of the quality parameters of a Mobile Telephony Communications Base Station such as signal level, SRN signal and signal quality and to perform a comparative analysis with what is established by the International Telecommunications Union (ITU), in its recommendation E-800. In addition, the level of relationship between the quality parameters (signal level, SRN signal and signal quality) was established, with the resolution of detecting that they comply with the ITU, this is confirmed by the optimal mobile phone coverage in rural areas, and that these have accessibility to the entire mobile phone network nationwide.

1. Introduction

Within the framework of the information society, the level of penetration of goods and services linked to mobile communications is undoubtedly the starting point for promoting public policies that promote digital connectivity. However, there are still large gaps that exist in both extension (access) and depth (quality of access) in rural areas throughout the world. [1]

According to the report on the state of broadband in Latin America and the Caribbean 2016, released by the Economic Commission for Latin America and the Caribbean (ECLAC), the number of internet users grew from 2000 to 2015 by 10, 6%, however, the problems regarding connection speeds and service access between urban and rural areas still make a big difference. [2]

The choice of technology to provide Internet connectivity services in general is varied, which is why, of all known technologies, very few or almost none is applicable to provide Internet connectivity services in rural areas, without However, due to its ability to access and reach, mobile phone technology easily adapts to the needs of the rural population. [3]

In recent years, there has been a rapid and continuous evolution of technologies in the field of mobile communications. [4] Mobile devices incorporate more and better services, and in addition to improving the quality of communications, it makes possible that many services are available to us. [5]

Given this increase in the demand for telecommunications services, with an increasingly demanding user profile in terms of mobility, availability, bandwidth or autonomy, the efforts of the operators to design and implement an infrastructure, capable of Support this increase in traffic level. [6]

Mobile connectivity has become a fundamental part of our lives, since they have become a determining technology in the way the user interprets the exchange of information. [7] The optimization of mobile telephony infrastructure is a frequent and necessary practice in operational telecommunications networks, in

* Omar Freddy Chamorro Atalaya. Jr. Los Damascos 986. Los Olivos, Lima, Perú, 968053089 & omar_chamorro1@hotmail.com

* Omar Chamorro-Atalaya, Email: omar_chamorro1@hotmail.com

all countries of the world, with the purpose of increasing the efficiency of its operation or performance to improve its quality of service. [8]

LTE technology is designed to work as a system with the ability to significantly improve user experiences with total mobility, since it uses the Internet Protocol (IP) to perform any type of end-to-end data traffic with good quality of service and in the same way the voice traffic, which allows a better integration with other multimedia services. [9]

The Peruvian mobile market is dynamic and growing, one could say that in general the user rate is one cell phone per inhabitant, as of March 2016, there were 30 million 98 thousand 298 active lines; The quality of service is the Achilles' heel of the Peruvian market operators, dissatisfaction in general reaches 36% of users. [10] Therefore, the priority of being able to count on the high availability of the service has motivated to evolve technologically in physical equipment that allows to carry effective telephony with quality of service. [11]

Given the mobility of users and the number of calls made daily, maintaining the quality of mobile voice service becomes a task that requires continuous monitoring and detection of areas where the service can be degraded. For this, the ITU (International Telecommunications Union), in recommendation E.800, has defined a series of network parameters as indicators of service quality. [12]

Although the predominant technology in the world today is LTE, even 2G networks are used mostly in rural areas or with very basic user profiles. [13] In rural areas there is a problem that is the densification and optimization of the 4G LTE network, which affects the quality of service offered by the network. [14]

In Peru, the deployment of 4G LTE technology has been carried out slowly and not entirely in some cities; this due to economic, social due and even cultural factors. [15] However, despite the important deployment of communications base station implementation by operators, according to a study conducted by OSIPTEL (Supervisory Body for Private Investment in Telecommunications), it is estimated that by 2021 there would be a gap of more than 15,200 stations to be installed. [16]

In addition, in the era of globalization that we currently live, the Internet has been essential in the communication of people worldwide, up to that point that sounds crazy without having this service in homes, businesses, companies, schools and universities. Thus causing a massive dependence and magnification of the service nationwide. That is where the main problem lies, since telecommunications companies focus their resources on the large cities located on the coastal coast and neglect the rest of the country, especially in rural areas where Internet access is almost non-existent or of poor quality. And from this problem was born the idea of analyzing the quality parameters of the implementation of a 4G-LTE Communications Base Station, installed in a Rural Area of Peru.

In this sense, this article describes the results obtained as part of the process of evaluating the quality parameters of a communications base station, implemented in a rural area of Peru. For this analysis the KPI (Key performance Indicator) defined for 4G LTE Technology was used.

2. Methodology

2.1. Research level

The research level is of the Descriptive-Correlational type, since it is sought to specify the characteristics of the quality parameters of a Mobile Telephony Communications Base Station such as signal level, signal SRN and signal quality and perform a comparative analysis with established by the International Telecommunications Union (ITU), in its recommendation E-800. [17]

And correlational, since, the relationship between the quality parameters (signal level, signal SRN and signal Quality) is determined, in order to verify that they comply with the ITU, thus guaranteeing optimal mobile phone coverage in the rural areas, and that these have accessibility to the entire mobile phone network nationwide. [18]

2.2. Population and sample

The unit of analysis is each of the elements that constitute the population and therefore the sample [17]; In that sense, my only unit of analysis will be Communications Base Station with 4G-LTE technology, whose implementation characteristics are shown below. The following table details the characteristics that the 700MHz band antennas possess, as well as characteristics such as Azimuth (inclination angle), the type of technology, Cell ID and the Band or frequency.

Table 1. Characteristics of the antennas in the 700 MHz band

Site: IC00243			
Name:	4GIC24371	4GIC24372	4GIC24373
Azimuth:	20	150	275
Technology:	LTE	LTE	LTE
Cell ID:	4	5	6
Band or Freq.:	700	700	700

As seen in the previous table there are three antennas, each one has three angles of inclination of 20 °, 150 ° and 275 °, the technology is LTE with 700MHz band. The following table shows the characteristics of the antennas in the 2100MHz band.

Table 2. Characteristics of the antennas in the 2100 MHz band

Site: IC00243			
Name:	4GIC24341	4GIC24342	4GIC24343
Azimuth:	20	150	275
Technology:	LTE	LTE	LTE
Cell ID:	1	2	3
Band or Freq.:	2100	2100	2100

As seen in the previous table there are three antennas, each one has three angles of inclination of 20 °, 150 ° and 275 °, the technology is LTE with a band of 2100MHz.

The following figure shows the Topology of connection of communication equipment that will be considered at the time of the implementation of the transmitting base station, with 4G-LTE technology.

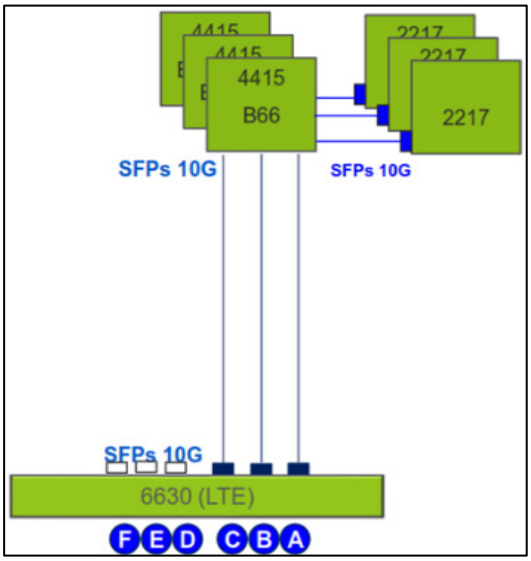


Figure 1: Equipment connection topology

In the previous figure, it is observed that a Baseband 6630 communication device will be used; which will support 4G technology (6630 LTE); in turn, the Baseband 6630 equipment will be connected to 3 RRU 4415 radio equipment (4G AWS: 1900MHz / 2100mHz) and 3 RRU 2217 radio equipment (4G-LTE: 700 MHz); The term RRU means "Remote Radio Unit".

Another important point that is convenient to detail are the characteristics of the radio equipment, as well as the antennas to be used, although it is true in the previous point the equipment was listed, then they detail certain particularities of the equipment used in the implementation process, such as frequency of operation, quantity of equipment and manufacturing brand.

Table 3. Equipment and antennas to be used in the implementation process 4G-LTE

Equipment	Quantity	Technology to install
Radio equipment to install	3	RRU 2217
Baseband to install	1	RBS 6630
Antenna to install	1	HUAWEI ASI4517R3v06
Rectifier and battery cabinet	1	Rectifier ZTE
Equipment cabinet	1	Equipment cabinet ZTE EX2
Others	9	Power cable 30m
	6	Fiber optic cable 50m
	5	Grounding rods
	25 meters	Cable GND

2.3. Data collection instrument

The instrument used in the investigation was the technical measurement reports, which details the characteristics of the quality parameters of a Mobile Telephone Communications Base Station such as signal level, signal SRN and signal quality.

3. Results

In order to determine the optimum quality parameters of the communications base station, the signal level, signal SRN and signal quality, on each carrier of the communications base station, was analyzed using the Smart Rollout Support program; These results are compared with what is established by the International Telecommunications Union (ITU), in its recommendation E-800A.

This entity indicates that the KPIs for 4G are the following: [19]

- Signal level: It is the average power received at the terminal. It is a measure of the signal and serves as the main indicator of coverage. An acceptable value of this measure is greater than -100dBm.
- Signal SRN: It is the ratio of power between the reference signal and interference plus noise. An acceptable value is greater than or equal to 20dB.
- Signal Quality: It is the relationship between the signal and the interference. An acceptable value of this measure is greater than -15dB.

Next, the following figure shows the results obtained from the 4GIC24371 antenna operational test with azimuth 20° with 700MHz carrier.

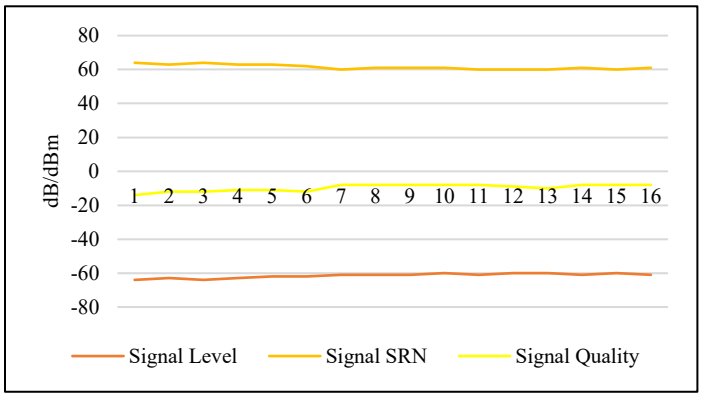


Figure 4: Antenna operational test 4GIC24371 20°

The previous figure shows that the average signal level is -61,375; The signal SRN is 61,375 and the signal quality is -9,5625; comparing these results with those described above it can be indicated that the antenna 4GIC24371 at an angle of inclination 20 °; It meets the quality parameters.

Next, the following figure shows the results obtained from the 4GIC24372 antenna operational test with azimuth 150 ° with 700MHz carrier.

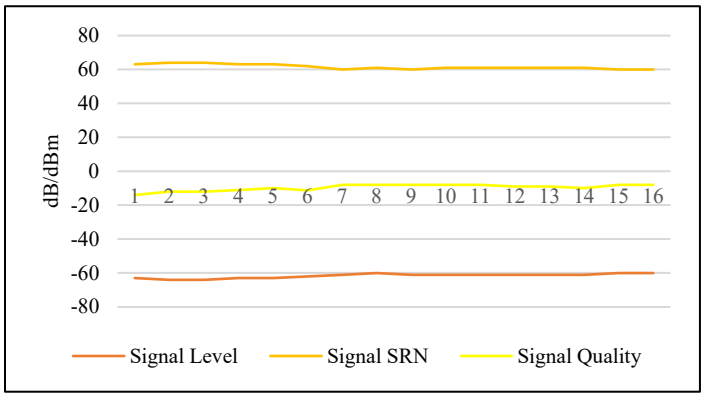


Figure 5: Antenna operational test 4GIC24372 150°

The previous figure shows that the average signal level is -61.5; The signal SRN is 61.5 and the signal quality is -9.6875; comparing these results with those described above it can be indicated that the antenna 4GIC24372 at an angle of inclination 150 °; It meets the quality parameters.

Next, the following figure shows the results obtained from the 4GIC24373 antenna operational test with azimuth 275 ° with 700MHz carrier.

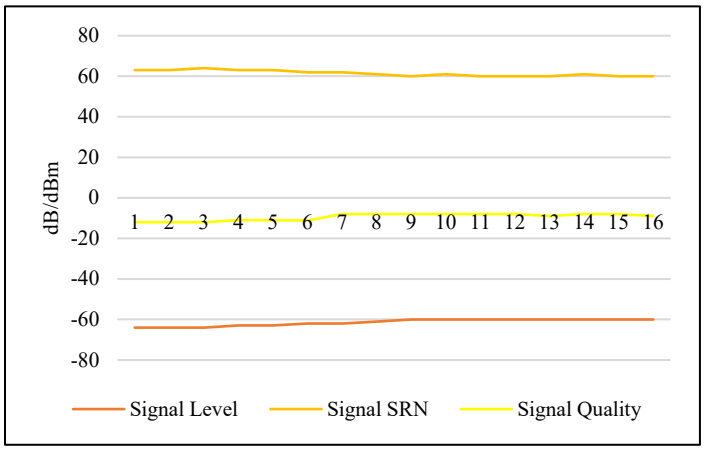


Figure 6: Antenna operational test 4GIC24373 275°

The previous figure shows that the average signal level is -61.5; The signal SRN is 61.5 and the signal quality is -9.6875; comparing these results with those described above it can be indicated that the antenna 4GIC24372 at an angle of inclination 150 °; It meets the quality parameters.

Next, the following figure shows the results obtained from the 4GIC24373 antenna operational test with azimuth 275 ° with 700MHz carrier.

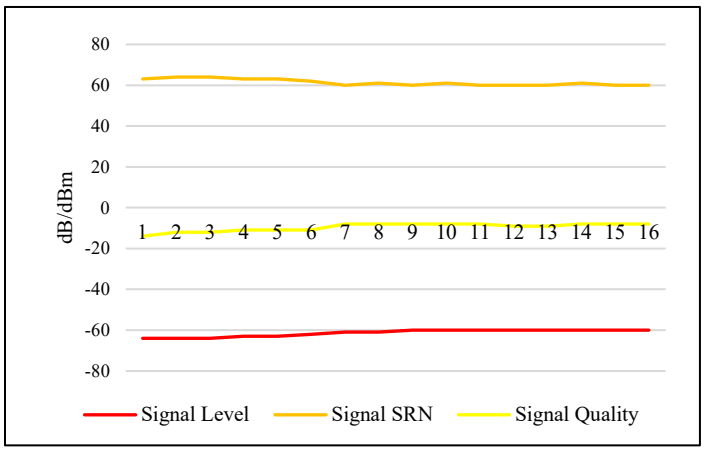


Figure 7: Antenna operational test 4GIC24341 20°

The previous figure shows that the average signal level is -61.4375; The signal SRN is 61,375 and the signal quality is -9,575; comparing these results with those described above it can be indicated that the antenna 4GIC24341 at an angle of inclination 20 °; It meets the quality parameters.

Next, the following figure shows the results obtained from the 4GIC24342 antenna operation test with azimuth 150 ° with 2100MHz carrier.

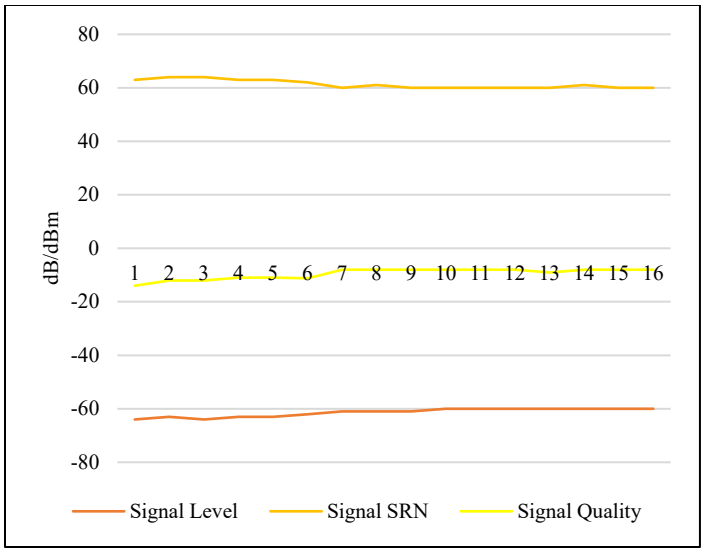


Figure 8: Antenna operational test 4GIC24342 150°

The previous figure shows that the average signal level is -61.4375; The signal SRN is 61.4375 and the signal quality is -9.45; comparing these results with those described above it can be indicated that the antenna 4GIC24342 at an angle of inclination 150 °; It meets the quality parameters.

Next, the following figure shows the results obtained from the 4GIC24343 antenna operational test with azimuth 275 ° with 2100MHz carrier.

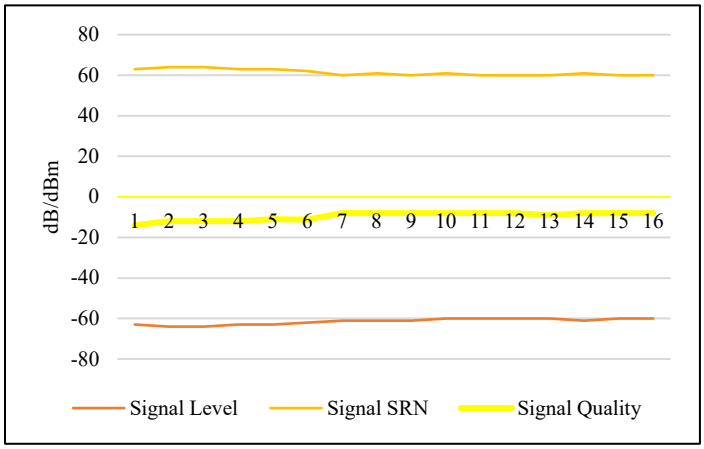


Figure 9: Antenna operational test 4GIC24343 275°

The previous figure shows that the average signal level is -61,375; The signal SRN is 61.3125 and the signal quality is -9.5125; comparing these results with those described above it can be indicated that the antenna 4GIC24343 at an angle of inclination 275 °; It meets the quality parameters.

4. Statistical analysis

The following table shows the result of the Pearson correlation test for the first carrier of the Communications Base Station; which will determine if there is an association between the quality parameters in analysis, such as signal level, signal SRN and signal quality, measured in the Mobile Telephony Communication Base Station. This test was performed using the SPSS statistical program.

Table 4. Pearson's correlation of the parameters of the first carrier

		Signal Level	Signal SRN	Signal Quality
Signal Level 4GIC24371 20°	Pearson's correlation	1	,936**	,914**
	Sig. (bilateral)		,000	,000
Signal SRN 4GIC24371 20°	Pearson's correlation	,936**	1	,869**
	Sig. (bilateral)	,000		,000
Signal Quality 4GIC24371 20°	Pearson's correlation	,914**	,869**	1
	Sig. (bilateral)	,000	,000	
Signal Level 4GIC24372 150°	Pearson's correlation	1	,935**	,832**
	Sig. (bilateral)		,000	,000
Signal SRN 4GIC24372 150°	Pearson's correlation	,935**		,846**
	Sig. (bilateral)	,000		,000
Signal Quality 4GIC24372 150°	Pearson's correlation	,832**	,846**	
	Sig. (bilateral)	,000	,000	
Signal Level 4GIC24373 275°	Pearson's correlation	1	,950**	,845**
	Sig. (bilateral)		,000	,000
Signal SRN 4GIC24373 275°	Pearson's correlation	,950**		,854**
	Sig. (bilateral)	,000		,000
Signal Quality 4GIC24373 275°	Pearson's correlation	,845**	,854**	
	Sig. (bilateral)	,000	,000	

As Pearson's correlation coefficient is shown in the previous table, it is interpreted as follows, for each indicator. [17]

- Signal Level - Signal SRN (4GIC24371 20 °): There is a very high significant ratio of 0.936 between the Signal Level and the Signal SRN.
- Signal SRN - Signal Quality (4GIC24371 20 °): There is a significant high ratio of 0.869 between the Signal SRN and the Signal Quality.
- Signal Quality - Signal Level (4GIC24371 20 °): There is a very high significant 0.914 ratio between Signal Quality with Signal Level.
- Signal Level - Signal SRN (4GIC24372 150 °): There is a very high significant ratio of 0.935 between the Signal Level and the Signal SRN.
- Signal SRN - Signal Quality (4GIC24372 150 °): There is a significant high ratio of 0.846 between the Signal SRN and the Signal Quality.
- Signal Quality - Signal Level (4GIC24372 150 °): There is a significant high ratio of 0.832 between Signal Quality and Signal Level.
- Signal Level - Signal SRN (4GIC24373 275 °): There is a very high significant ratio of 0.950 between the Signal Level and the Signal SRN.
- Signal SRN - Signal Quality (4GIC24373 275 °): There is a significant high ratio of 0.854 between the Signal SRN and the Signal Quality.
- Signal Quality - Signal Level (4GIC24373 275 °): There is a significant high ratio of 0.845 between Signal Quality with Signal Level.

The following table shows the result of the Pearson correlation test for the second carrier of the Communications Base Station.

Table 5. Pearson's correlation of the parameters of the second carrier

		Signal Level	Signal SRN	Signal Quality
Signal Level 4GIC24341 20°	Pearson's correlation	1	,958**	,880**
	Sig. (bilateral)		,000	,000
Signal SRN 4GIC24341 20°	Pearson's correlation	,958**	1	,890**
	Sig. (bilateral)	,000		,000
Signal Quality 4GIC24341 20°	Pearson's correlation	,880**	,890**	1
	Sig. (bilateral)	,000	,000	
Signal Level 4GIC24342 150°	Pearson's correlation	1	,958**	,903**
	Sig. (bilateral)		,000	,000
Signal SRN 4GIC24342 150°	Pearson's correlation	,958**		,850**
	Sig. (bilateral)	,000		,000
Signal Quality 4GIC24342 150°	Pearson's correlation	,903**	,850**	
	Sig. (bilateral)	,000	,000	
Signal Level 4GIC24343 275°	Pearson's correlation	1	,933**	,926**
	Sig. (bilateral)		,000	,000
Signal SRN 4GIC24343 275°	Pearson's correlation	,933**		,895**
	Sig. (bilateral)	,000		,000
Signal Quality 4GIC24343 275°	Pearson's correlation	,926**	,895**	
	Sig. (bilateral)	,000	,000	

As Pearson's correlation coefficient is shown in the previous table, it is interpreted as follows, for each indicator. [17]

- Signal Level - Signal SRN (4GIC24341 20 °): There is a very high significant ratio of 0.958 between the Signal Level and the Signal SRN.
- Signal SRN - Signal Quality (4GIC24341 20 °): There is a significant high ratio of 0.890 between the Signal SRN and the Signal Quality.
- Signal Quality - Signal Level (4GIC24341 20 °): There is a significant high ratio of 0.880 between Signal Quality with Signal Level.
- Signal Level - Signal SRN (4GIC24342 150 °): There is a very high significant ratio of 0.958 between the Signal Level and the Signal SRN.
- Signal SRN - Signal Quality (4GIC24342 150 °): There is a significant high ratio of 0.850 between the Signal SRN and the Signal Quality.
- Signal Quality - Signal Level (4GIC24342 150 °): There is a very high significant 0.903 ratio between Signal Quality with Signal Level.
- Signal Level - Signal SRN (4GIC24343 275 °): There is a very high significant ratio of 0.933 between the Signal Level and the Signal SRN.

- Signal SRN - Signal Quality (4GIC24343 275 °): There is a significant high ratio of 0.895 between the Signal SRN and the Signal Quality.
- Signal Quality - Signal Level (4GIC24343 275 °): There is a very high significant ratio of 0.926 between Signal Quality with Signal Level.

5. Discussion

In relation to the results of this research, the following discussions are held below:

In relation to the quality parameters, which present optimal operating values, these results are similar to those carried out at the Real Plaza Santa Clara Shopping Center, where it is observed that when implementing 4G technology; 100% of the measures taken have levels greater than -95dBm. Depending on the allowed range (RSRP> - 100dBm), the values obtained are entirely within this range. This ensures optimal 4G coverage within the CC with levels that exceed those obtained in the simulations. [19]

In relation to the association of the quality parameters, it can be affirmed that the results comply with what was indicated by the investigation of the Canary Islands Government, which indicates that Signal Quality is also measured by the Signal SRN relationship, the higher the value of the Signal SRN higher quality will have the same. [20]

Optimal signal coverage is obtained with the results. This is supported by the KPI levels established by the International Telecommunications Union (ITU), in its recommendation E-800A. [19]

6. Conclusions

It is concluded that the results obtained in relation to the transmission quality parameters of the Telephone Base Station, signal level, signal SRN and signal Quality, comply with the provisions of the International Telecommunication Union (ITU), in its E-800 recommendation.

It is concluded that there is a very high significant relationship between the Signal Level and the Signal SRN; there is a significant high relationship between the Signal SRN with the Signal Quality and there is a high significant relationship between the Signal Quality with Signal Level.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors wish to recognize and thank the National Technological University of Lima South Lima for their support of this investigation and

References

[1] Silva, A. (2018). A regional look at access and possession of information and communications technologies - TIC, based on censuses. CEPAL Magazine, 167-171.
[2] Radicelli, C., Pomboza, M., & Cepeda, L. (2017). Connectivity to the Internet in rural areas using DTT (DVB-RCT2) technologies, or mobile telephony (4G-LTE) •. DYNA, 319-324.
[3] Ibarra, L. (2016). Technological Characteristics of 3G and 4G LTE mobile systems: Current situation in Ecuador. Catholic University of Ecuador, Guayaquil.
[4] Acevedo, R. (2017). Automatic planning of parameters in LTE Networks through Graph theory. University of Malaga, Malaga.
[5] Gracía, F. (2017). Prototype of a 4G Base Station, using Open Air Interface. University of Granada, Granada.
[6] Liu, H. (2016). Evaluation of a Software based LTE Base station. Higher Technical School of Telecommunications Engineering of Barcelona, Barcelona.
[7] Jiménez, J., & Ochovo, J. (2017). Comparative Evaluation of Mobile Networks. Polytechnic University of Madrid, Madrid.
[8] Limonchi, C. (2016). Optimization proposal for the Access Network in the 1900 MHz Band of the Mobile Bitel Operator to Improve Coverage in the Chiclayo Zone 2016. Juan Mejía Baca University, Chiclayo.
[9] Alvarez, R. (2016). Improvement of the Quality of Service in 4G Networks Applying Offloading Technology. Santa María Catholic University, Arequipa.
[10] Guerrero, C. (2016). Determination of Satisfaction Indexes of cell phone customers to improve the quality of service in the Province of Huancayo. National University of Central Peru, Huancayo.
[11] Jara, C. (2017). Technological Renewal GSM / UMTS / LTE SWAP project 2014-2015. Private University of Tacna, Tacna.
[12] Marín, M. (2017). System for Measuring and Analysis of Call Establishment Time. University of Chile, Santiago de Chile.
[13] Córdova, R. (2019). Design of a scalable LTE network for the transmission of video applied to citizen security. Peruvian University of Applied Sciences, Lima.
[14] Chino, E. (2018). Design of an advanced 4G LTE Network improving the Access Network Identification through Heterogeneous Cell deployment for the City of Juliaca. National University of the Altiplano, Juliaca.
[15] Puel, D. (2017). Design of a 4G-LTE Outdoor Res from América Movil Perú SAC, for the Province of Sechura. National University of Piura, Piura.
[16] More, J., Trelles, J., & Pacheco, L. (2017). Supervisory Body of Private Investment in Telecommunications.
[17] Hernández, R. (2018). Investigation Methodology Mexico: Mcgraw Hill
[18] Arias, F. (2016). Research project Caracas: Episteme
[19] Chancasana, M. (2016). Design of a 4G-LTE Indoor Network for the Real Plaza Santa Clara Shopping Center. Pontifical Catholic University of Peru
[20] Government of the Canary Islands (2016). Telecommunications and computer science. Spain

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 77-85 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Routing Protocols for VANETs: A Taxonomy, Evaluation and Analysis

Amina Bengag*, Asmae Bengag, Mohamed Elboukhari

MATSI lab, ESTO, Mohammed First University, Oujda, Morocco

ARTICLE INFO

Article history:
Received: 21 September, 2019
Accepted: 27 December, 2019
Online: 15 January, 2020

Keywords:
VANETs
Routing protocols
SUMO
NS-3
OSM
Urban scenario
Greedy forwarding
Angle direction

ABSTRACT

VANET as a subclass of MANET is composed of a set of vehicles equipped with wireless transceivers, to build dynamic networks without the need of any pre-existing infrastructure. Over the last few decades, the area of routing protocols in VANETs has been extensively studied. Nevertheless, this area remains even more challenging due to some features of VANETs, such as the high speed of vehicles, the often-disconnected links and the particular mobility pattern. Routing protocols in VANETs could be splitted into four categories: topology, position, multicast and broadcast-based routing protocols. In this paper, we provide a novel detailed taxonomy of routing protocols in VANETs then we present the advantages and drawbacks of each category. Moreover, we clear up the techniques adopted by each of the most popular routing protocols based on the vehicles' position and the topology of the networks. To explore the strengths and weaknesses of each routing protocols, basing on their suitability for VANET, we implemented them by using SUMO and NS3 as simulation tools applied on a real street map of Oujda city. We have extracted the used map from Open Street Map (OSM). Finally, we present our future works used for optimizing the greedy forwarding technique that is adopted by some position-based routing protocols in VANETs. Our suggested technique is based on the angle direction and three other important parameters of the relaying node.

1. Introduction

In this paper, we review and provide an extension version of previous work originally presented in the 5th International IEEE Congress on Information Sciences and Technology (IEEECiST'18) [1], in addition to some new enhancements. It is about routing protocols in Vehicular Ad-hoc Networks.

Vehicular networks have become the heart of the Intelligent Transport Systems or ITS in short, to improve the road traffic safety and to make a comfortable driving environment for drivers and passengers. In order to establish a communication in VANET, there is no need to any pre-existing infrastructure; this kind of networks consist of a set of vehicles containing wireless transceivers, named On-Board Unit (OBU). These OBUs help vehicles to exchange information between each other.

Compared to other sub-category of Mobile Ad-Hoc Networks, vehicular networks have many exclusive and specific features, such as diverse communication environments. Actually, in the rural environment, the issue of sparse connectivity is high and the traffic condition is simple; while for the city environment, the density and the number of obstacles (such as building trees and others) are high, which make the traffic condition very complex. Besides, the high speed of vehicles could be the main cause of the high dynamic topology and the regularly link breakage issue as the second and third features of VANETS. The last one is the variable network density, which is caused by the fact that the traffic density varies. This later depends on the type of environments and the moment of the day, which is very high in rush hour and medium or low in the other time.

Due to the above-mentioned features of vehicular networks, the proposition of an effective routing technique remains even more challenging to overcome. In this paper, we give our proposed taxonomy of routing techniques in vehicular networks [2] into routing based on network's topology, vehicle's position, multicast technique and broadcast technique as is shown in Figure 1.

The main objective behind the presentation of this taxonomy is to specify the advantages and disadvantages of each routing protocols' class, then declare the most convenient categories to VANETs. Furthermore, this declaration will be proved by a comparison study that is provided under two simulation tools. We also introduce our future work in which we suggest a new strategy

*Corresponding Author: Amina BENGAG, bengag.amina@gmail.com

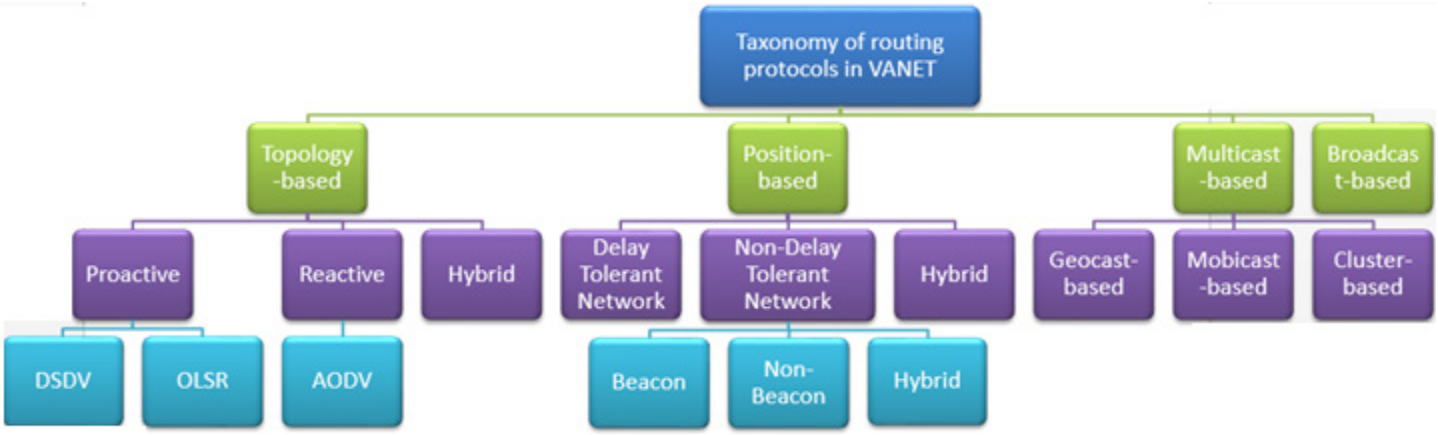


Figure 1 Taxonomy of VANETS routing protocols

to enhance the high-performed routing protocols, which belong to the category based on vehicles' position. This new strategy proved through theoretical analysis, is based on four parameters like velocity, the direction of vehicles, the density and the vehicles' position.

The rest of this paper is organized as follows: in section II, we outline the related works. In section III, we provide a detailed categorization of VANET's routing protocols and discuss their benefits and issues [2]. Section IV discusses in detail a comparison study of some routing techniques, based on the network topology and others on the physical position of vehicles by utilizing two simulation tools like NS3 as a network simulator and SUMO [1] as a road traffic simulator. Then, in section V, we introduce our current work and the proposed technique to optimize the Greedy forwarding strategy in GPSR protocol. Finally, in section VI, we conclude the paper and we present some our future works.

2. Previous work

To select the best-suited protocol to Vehicular Networks, many studies have been conducted to evaluate and examine the performance of certain routing mechanism based on simulation tools. In Table 1, we present some of the previous works related to the mentioned above study. It could be concluded from the table that all works is based on ns-2 or ns-3 as networks simulators, while some of them use also a specific road traffic simulators such as SUMO. However, in this work we provide a new study distinguish from those cited in Table I in many sides.

Basically, our work in this paper, focused on modeling the real details of VANETs network movement to examine the performance of certain routing protocols: DSDV, AODV, GPSR, OLSR, and GPCR. We extracted a part of a real road map from open street map (OSM) of a large street in Oujda city (Hay al-Quds), instead of considering a grid map or a random map. This map was made to be as an input file in SUMO in order to create mobility and traffic files; those files will be used as an entry in NS3 for further analysis of network performance.

3. Taxonomy of Routing Protocols in VANETs

The regular link breakage issue of vehicular networks caused by the speed of vehicles that lead to the high dynamic topology in these networks; the mentioned problems make the task of transmitting data packet in vehicular networks a very challenging one. Mainly, routing protocols in VANET can be classified into four categories [2] for V2V communication: based on the network's topology, based on the vehicles' position, based on multicast technique and finally routing based on broadcast technique.

Table 1 Related Works

Title and Reference	Simulators	Routing Protocols	Results
"Performance evaluation of olsr and aodv protocols in a vanet crossroad scenario"[3]	NS-3 CAVENET	OLSR, AODV	OLSR is the best one
"Comparative analysis of various routing protocols in vanet"[4]	NS-2.34	AOMDV, AODV, DSDV, DSR	The best performance is gotten in case of DSR for throughput and E2ED. Whereas, AOMDV and AODV perform better in terms of PDR
"Performance of routing protocols for vanets: A realistic analysis format" [5]	NS-3 and VANETMobiSim	AODV, OLSR	OLSR has the best performance

"Scenario based performance analysis of aodv and gpsr routing protocols in a vanet"[6]	NS-2.35 and VANETMobi Sim	AODV, GPSR	AODV outperforms GPSR in case of PDR metric and GPSR behaves better for E2ED
"Performance comparison of position based routing protocols for vanets"[7]	NS-2 and VANETMobi Sim	GPSR, GPCR	GPCR is the best one
"Performance Evaluation of OLSR and AODV in VANET Cloud Computing Using Fading Model with SUMO and NS3"[8]	NS-3 and SUMO	AODV, OLSR	Almost the same performance for both protocols

3.1. Routing based on the network topology

In this category, picking an itinerary between the source and the target node is based the existing link information in the network. The collected information could be done before any need of routing by the vehicle by utilizing proactive technique that is known by the table-driven technique. Moreover, for reactive technique, the information is searched when we need to transmit data; this procedure is also known by on-demand technique, in which the step of finding or maintaining an itinerary from source vehicle to destination is essential before any transmission.

In proactive technique, nodes collect and maintain updated routing data beforehand by periodically distributing routing tables throughout the network to all the nodes in its transmission range. The control messages in proactive routing are broadcasted continuously even if there is no data flow to send. Hence, the routing overhead is very high and the PDR is low consequently. The benefit of this category of routing protocols is that nodes can found routing information easily. DSDV, OLSR are two examples of proactive routing protocols.

In reactive technique, nodes start searching for a route only if they need to transmit data. Hence, they flood the network with Route Request packets (RREQ). AODV and DSR are two examples from this category. There is also one more type called hybrid protocol that associates reactive and proactive techniques, which make them more efficient. Zone Routing protocol (ZRP) is an example of such routing technique [9].

3.2. Routing based on vehicle's Position [10]

This class of routing protocols is also known as geographic protocols. They require the availability of a positioning system that provide the physical location of the participating vehicles in the relay selection process. The utility of this physical position information is identifying and making decision to select the next hop to route packet to the last destination. This technique reduces the routing overhead issue because it does not require any routing tables or exchange any information between them to establish routes. Here, it is not essential to identify the whole route to deliver the data packets that reduce the end-to-end delay. Based on several previous works, we can divide the routing based on vehicle's position into three sub-categories: routing adapted to Delay Tolerant Network (DTN), routing adapted to Non-Delay Tolerant Network (Non-DTN) and hybrid position-based routing.

Routing protocols for DTN: this category of routing techniques is considered as an efficient one for networks that have some specific features, such as high link breakage, limited bandwidth, power constraint and large scale. Those problems could be reduced by using a carry-and-forward technique, in which the transmitted vehicle stores the packet until it finds another vehicle moves into its vicinity. Vehicle-Assisted Data Delivery (VADD) [11] is an example of such routing technique.

Routing for Non-DTN: those protocols are characterized by the regularly link breakage issue that cause the lack of connectivity and a lack of instant end-to-end routes. The greedy approach is the process adopted by this routing to look for the next relaying vehicle. Actually, by utilizing this strategy, the packet is transmitted to the nearest neighbor of the destination. While the transmission failure could happen in anytime especially when a node finds no neighbors nearer to the destination than itself. In non-DTN protocols, we can find three sub-class; routing based on beacon that use a periodic hello message, others are beaconless, and hybrid protocols. GPSR and GPCR are an example of such routing technique considered as beacon-based protocols.

Hybrid position-based Routing protocols: this category of routing techniques could be considered as a combination of the two previous mentioned techniques. The main advantages of this category is reducing the control overhead and resolving the frequently link breakage problem in VANET. GeoDTN+Nav [12] is an example of such routing technique.

3.3. Routing based on multicast technique [13]

In routing based on multicast technique, the data packet is forwarded from a single source to a specific group of vehicles by using multi-hop communication technique. This protocol could be divided into three categories: protocols based on the geocast technique, based on Mobicast technique and those based on clusters.

- Routing based on Clusters [14] here, the vehicles share the same features such as the direction of mobility with almost the same velocity can form a cluster, and in each one a node will be elected to manage the communication into each cluster and between those clusters. This node is called a cluster head (CH). Cluster based Routing (CBR) [15] is an example of this category.

- Geocast-based routing protocols [16] here, data is forwarded from a single vehicle as source to vehicles belonging to multicast group or zone of relevance (ZOR) based on the physical position of vehicles. One of the most known routing in this category is Inter-Vehicule Geocast (IVG) [17].
- Routing based on Mobicast technique [18] it is a kind of geocast-based protocol that are based in time as another factor in addition to the space. Actually, by using this method, data packets will be transmitted to all vehicles situated in certain geographic area at a specific time t so called ZORt. One of the most popular routing in this class is Mobicast routing protocol with carry and forward [18].

3.4. Routing based on broadcast technique [19]

In the broadcast technique, the data packets will be transmitted to reach the whole vehicles in the network. This technique is usually used in vehicular networks to share some important information between vehicles, such as road conditions, traffic, weather and emergency event. Based on several previous works [20], [21], we can divided this class of routing protocols into three categories as shown in Figure 2. These categories could have several common properties such as information type, information source, broadcast decision and orthogonal enhancements.

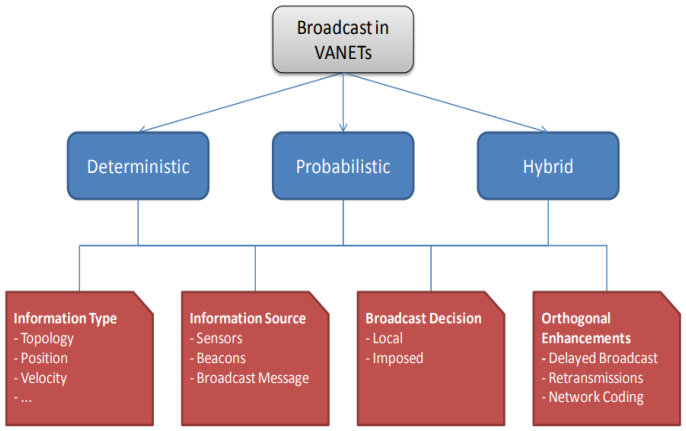
The three classes of routing protocols based on broadcast techniques are:

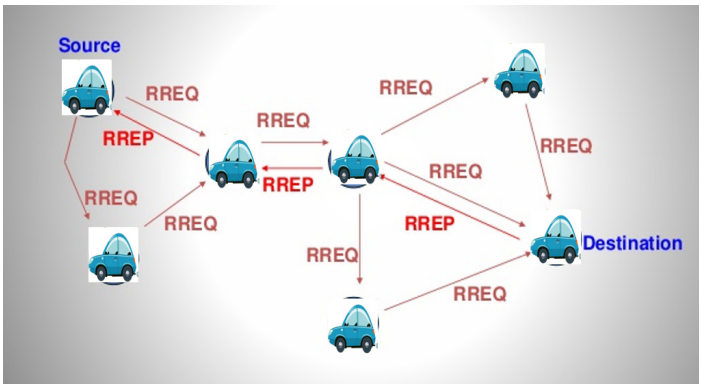
- The deterministic that is based on the topology of the network, the position and the velocity of the vehicles to rebroadcast the data packets.
- The probabilistic in which a vehicle will rebroadcast a data packet with a certain probability fixed in priory or adapted periodically.
- The hybrid broadcast that is a combination of deterministic and probabilistic techniques. In fact, the hybrid broadcast technique uses the deterministic technique to compute the most convenient broadcast probabilities.

Distribution Vehicular broadCAST (DV-CAST) [22], Distribution-Adaptive Distance with Channel Quality (DADCQ) protocol [23], Link-based Distributed Multi-hop Broadcast (LDMB) [24] and Optimized Position-based Gossiping (OPbG) protocol [25] are some examples of this category.

4. Comparison of routing based on network's topology and vehicle's position

This section reviews in detail the evaluated routing protocols that includes routing based on network's topology and others based on vehicle's position.





4.1. Routing based on network's topology

- AODV routing protocol. AODV [26] could be classified as unicast or multicast routing, the process adopted by this routing is that each paths are produced only on request. In fact, when a vehicle desires to transmit a packet by using the route discovery mechanism as Figure 3 shows, only the used routes are kept. In this case, the route discovery process causes significant delays before every data transmission to look for the path. AODV is also known by its capacity to diminish broadcasts, transmission latency and routing overhead. However, AODV suffers from the high end-to-end delay resulted by the route discovery process before every data transmission. Hence, the high E2ED is not suitable to vehicular networks in case of a crucial or dangerous information.
- OLSR routing protocol. In OLSR [27], each node constructs a global view of the network's topology by distributing after each specific time its routing table. In this routing technique, the concept of Multi point Relays (MPR) is utilized to relay control traffic. The MPR technique decreases the routing overhead issue by selecting only specific nodes to retransmit the control messages, which aids in optimizing the broadcast technique.

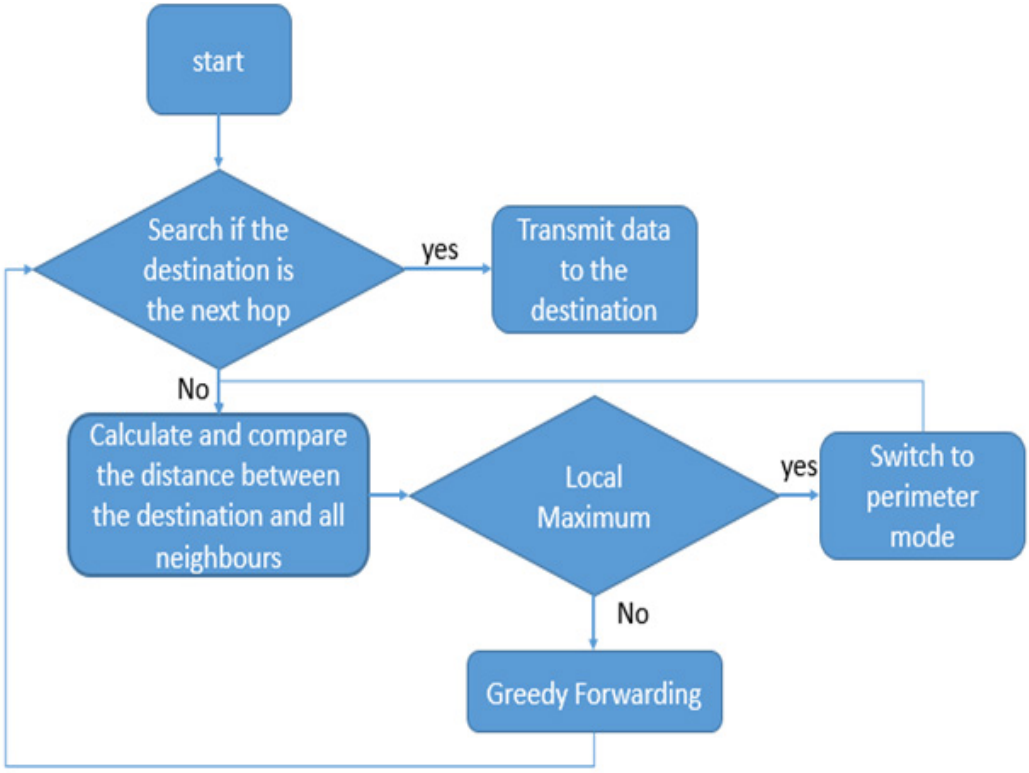


Figure 4 Flow diagram of route maintenance in GPSR

- DSDV routing protocol. In DSDV [28], [29], the update of routing tables of each nodes is done periodically to make available information about paths to each destination in the network at any time, even if the paths are at this time unused. In spite of the benefits of DSDV such as simplicity, loop free and no delay time caused by the discovery of the road technique. The DSDV still suffer from the high routing overhead caused by the periodic updates.

4.2. Routing based on vehicle's position

- GPSR [30] routing protocol. This routing protocol requires the availability of the physical location information of participating nodes using a location service like a GPS. In GPSR, the greedy forwarding (GF) technique is used by each relaying vehicle in order transmit data. In case of the local maximum problem, this technique fails; hence, the perimeter forwarding technique is applied to select another node through which a packet will travel. GPSR has many benefits: to transmit a data packet, a node necessities to memorize only the physical location of one hop neighbor. However, because of the features of high speed of vehicles, stale information of the neighbors' physical location are frequently contained in the sending nodes neighbor table. To be more clear, the strategy of GPSR is explained in Figure 4.
- GPCR routing protocol. Greedy Perimeter Coordinator Routing [31] protocol is an enhanced version of GPSR that uses a restricted greedy forwarding technique and a junction-based repair strategy in which routing decisions are made. Therefore, all forwarded packets must be sent to a vehicle located at an intersection or a junction; then this vehicle will transmit those packets to the closest neighbor of the recipient. GPCR has a great benefit that reduce the routing overhead problem and the end-to-end delay. Indeed, in the transmission process the election of the next hop is depends only on junction vehicles, which does not require any global or external information. Nevertheless, GPCR has some drawbacks that badly affect its performance. Actually, this technique assumes that there is at all times at least one vehicle at an intersection, which is not always the case.

4.3. Simulation and comparison

To simulate and evaluate the performance of the above-mentioned routing protocols in VANETs, it would be very expensive to set up a network to test certain criteria. To remedy this problem, we will use network and traffic simulators. In this project, we realize a scenario where we test the performance of routing protocols by successively increasing the number of vehicles in the network, while the other network parameters are fixed.

Basically, our work was evaluated by computing the rate of three metrics as the received packet number, the end-to-end delay and the routing overhead. To ensure this task, a simulation was carried out under the network simulator NS3 [32] version 25 in combination with the road traffic simulator SUMO 0.25 [33] by using the operating system Ubuntu 14.04. We have also chosen the Open Street Map (OSM) [34] to download a real-world urban map for a part of Oujda city (Hay al Quds) in Morocco; after that,

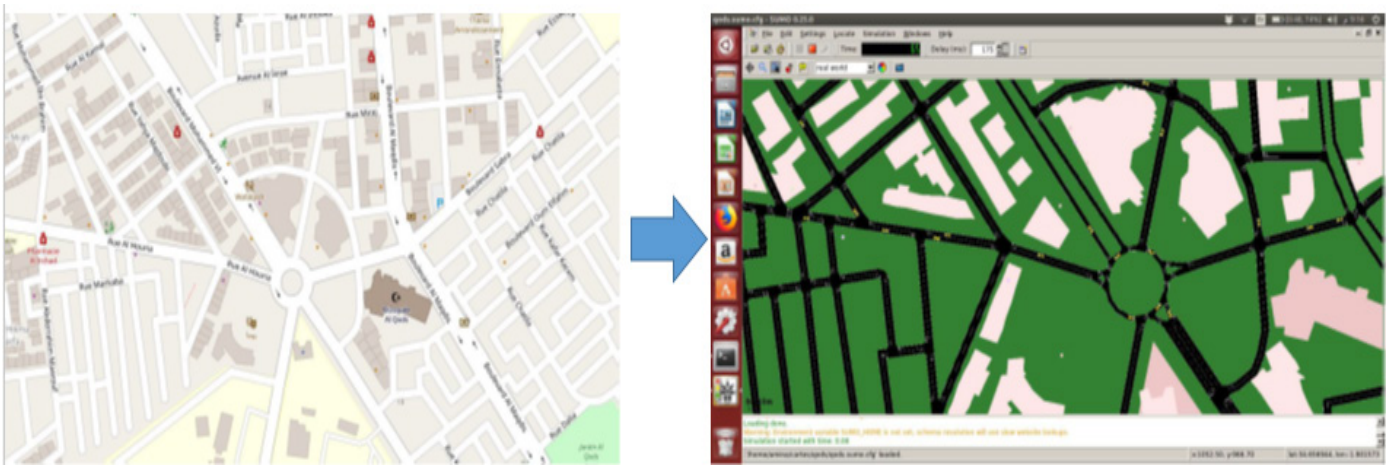


Figure 5 Hay Alquds Oujda map from OSM to SUMO

we have used SUMO to generate the traffic scenario as shown in Figure *5*.

In the used scenario, the number of nodes connected in the network varies. The simulation configuration is shown in Table *1*.

Table 1 Setting of experimental simulation parameters

Parameters	**Measures**
Routing Protocols	AODV, OLSR, DSDV, GPSR, GPCR
Simulation Object	Hay-ElQuds Oujda city
Number of nodes	20,30, 40, 50, 60, 70, 80, 90
Vehicles speed	Max : 20 m/s
Simulation time	100 s
Simulation area	1.7 * 1.5 km^2
Packet size	512 bytes
Data type	Constant Bit Rate (CBR)
Transport protocol	UDP
Propagation model	Two Ray Ground
Mac protocol	IEEE 802.11p
Channel	Channel/WirelessChannel
Network interface	Phy/WirelessPhy
Transmission range	145m
Transmission power	20 dBm

- Packet Delivery Ratio (PDR)
  According to the Figure 6, the results obtained confirm that DSDV and OLSR protocols have the best values of PDR. These values are increased when the vehicle's number increases, while they are reduced in case of AODV. Moreover, the curve stays stable For GPCR and GPSR protocols but with a moderate values of PDR. In fact, the huge number of radio obstacles in the proposed scenario are the main cause of this weak. Those obstacles increase the possibility of getting into the local maximum's trouble as explained in section IV. In addition, the PDR for GPSR and GPCR may be negatively affected because of the perimeter mode technique where the number of hops become high. For OLSR protocol, the good values obtained may be thanks to the use of Multi-Point Relay technique, which has a positive effect in case of dense and large networks.

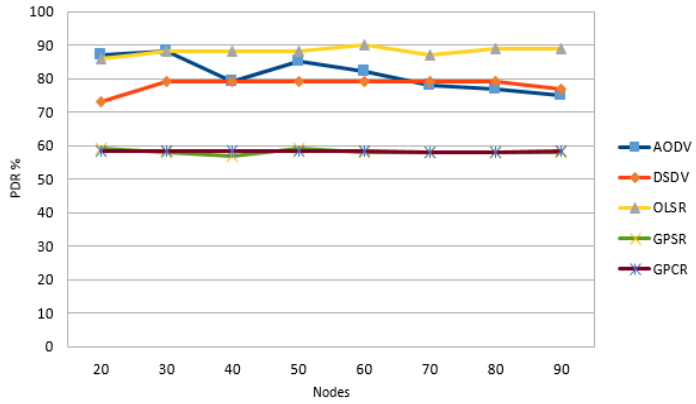


Figure 6 Packet Delivery Ratio Vs number of vehicles

- Average End-to-end Delay
  Average End-to-end Delay or E2ED in short denotes the required time for transmitting a packet from the source node to the destination per the number of successfully transported and received packets. The protocol that has an average E2ED as low as possible is the good and the preferred one. To calculate the E2ED we used the equation (2).

$$E2ED = \frac{\sum_1^n receptionTime - SendTime}{ReceivedPackets} \quad (2)$$

From the graph showed in Figure 7, we approve that the OLSR protocol has the maximum values followed by DSDV. Conversely, GPSR and GPCR have the lowest values, which are almost the same for all the number of nodes. The achieved results for GPSR and GPCR could be thanks to the use of the greedy forwarding technique. By using this technique, the number of hops are reduced so as the necessary time for transmitting data packets until the endpoint vehicle. It is also clear, that AODV has the best values in case of the E2ED compared to DSDV and OLSR for low node density. However, for the high node density, the E2ED for AODV are high because of the use of the route discovery technique that introduces additional delay.

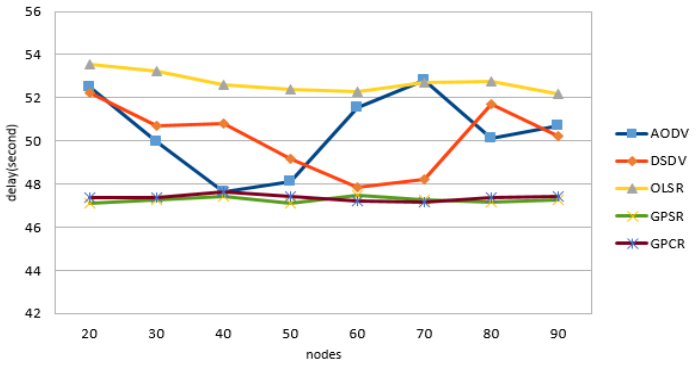


Figure 7 E2ED vs number of vehicles

- Throughput:
  The graph in Figure 8 proves that OLSR has the best and highest values of throughput up to 18kbps. In fact, the use of Multi point relay technique reduces the number of relaying nodes so as the routing overhead that helps in increasing the throughput. In case of DSDV and AODV, the throughput decreases as the density increases; this could be explained by the fact that DSDV broadcasts the entire routing table after fixed time interval, which introduces an additional overhead and affects its throughput. While GPSR and GPCR have the lowest throughput compared to the other three protocols.
- Routing Overhead:
  The results obtained in Figure 9 show that the best results are in case of GPSR and GPCR protocols that have the smallest values of routing overhead. This result could be explained by the fact that GPSR and GPCR are based on the physical location of the destination that is contained in the routed packet so that the retransmitted node should not need to use the location service for a second time, which reduces the overhead. The highest control overhead is for AODV, as it broadcasts a large number of control message packets to maintain the route.

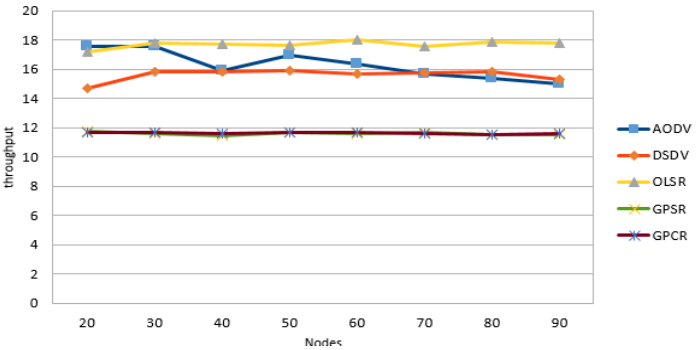


Figure 8 the throughput vs number of vehicles

5. Work in progress

In our proposed scenario, the performance evaluation above proves that OLSR has the best results in case of PDR and throughput. At the same time, GPSR and GPCR have the best results in case of overhead and E2ED. Indeed, the last two protocols are based on the destination's position that is carried in the packet, so the transferring vehicles do not need to use the location service for a second time and thus reduce the issue of routing overhead.

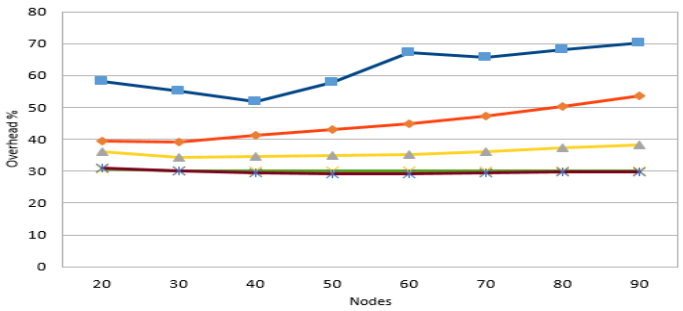


Figure 9 routing overhead vs number of vehicles

GPSR and GPCR have a moderate amount of PDR and Throughput. Hence, to optimize the ratio of the mentioned parameters, we proposed a new algorithm that improves the classical technique of greedy forwarding. This novel technique is used to select the best relaying vehicle candidate by taking into consideration four main metrics: the angle direction, the speed variation, the density of each nodes in addition to the distance between the next-hop candidate vehicle and the destination vehicle.

Our main contribution consists in suggesting a new greedy forwarding strategy based on a simple weighted function (WF), which uses the four last mentioned metrics. Then, an improved GPSR and GPCR protocols will be provided based on our proposed strategy. More specifically, the chosen of the best relaying node candidate is done by selecting the neighbor node whose header angle is closest to the destination node. Therefore, the connection between the source and the destination consists of a series of nodes that move approximately in the same direction. The main goal of this concept is to maintain the connection between the nodes as long as possible, in order to reduce the problem of link breakage. This is in contrast to the simple greedy forwarding, which only considers the distance between the next hop and the destination and does not take into account the connection lifetime. The new technique also could enhance the stability of the routing link, which mitigates the influence of high dynamic topology in VANETs networks.

In GPSR and GPCR, without considering moving direction of vehicles while looking for the next optimal hop, the routing techniques could produce an erroneous decision to transmit a data packet, which could result in a high packet loss issue. In addition, the regularly link breakage problem may also disturbs the quantity of delivered packets, particularly in highway environment where the speed is very high. Basically, when a node receives a Hello packet to up-to-date its neighbor's list, it computes the best next hop and then forwards packet. Thus, during this process, the selected neighbor could be outside the transmission range and the packet cannot reach it due to the high network's mobility. In order to solve those issues that lead to performances degradation, we

thought to use other parameters as mentioned above like the speed, the density of neighbors and the distance between all participating nodes and the destination during routing operation in addition to the moving direction. Furthermore, a simple weighted function (WF) will be applied on each relaying candidate vehicle, then the relaying candidate node that has the lowest value of the WF will be selected.

The proposed technique resolves mainly three problems in vehicular networks, which could be proved through theoretical analysis:

1. The problem of void area is resolved by taking into account the density (number of neighbors) of the selecting node. In Figure 10, we present the density parameter or the number of neighbors of a selected node. Basically, vehicle A has the vehicles B to F as neighbors because they belong to its transmission range.

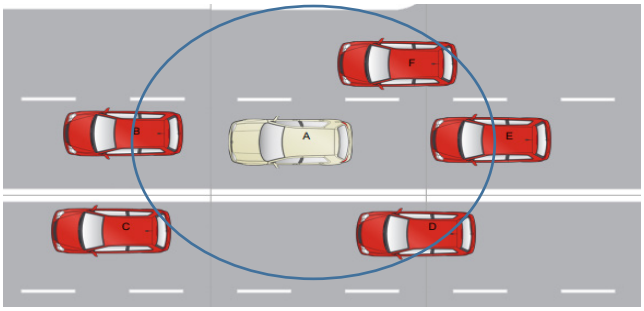


Figure 10 neighbors of the node A (density)

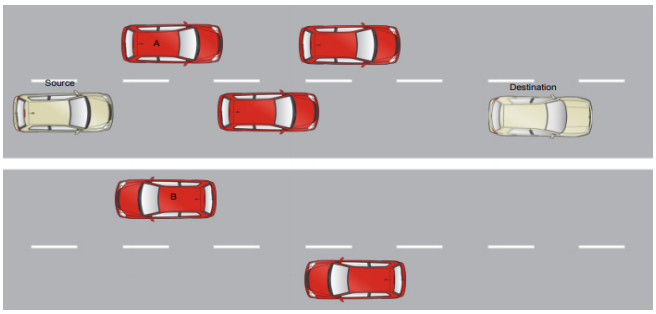


Figure 11 void area problem

Indeed, by using the proposed technique the node that has the big number of neighbors increases the probability to be chosen as the next hop. Figure 11 presents an example in which the source will choose A as the next hop because it has three neighbors while B has no neighbors.

2. The problem of direction of nodes in VANETs will be fixed by taking into account the angle direction calculated between the next hop and the destination node. As the Figure 12 shows, the source node will choose A as the next hop by applying our proposed technique because it has almost the same direction as the destination. The angle direction between the next hop and the destination is calculated by using our proposed formula (3).

$$\varphi_{id} = \cos^{-1}\frac{((iVelocity.x*dVelocity.x)+(iVelocity.y*dVelocity.y))}{(\sqrt{(iVelocity.x^2+dVelocity.x^2)}*\sqrt{(iVelocity.y^2+dVelocity.y^2)})} \quad (3)$$

Where iVelocity is the velocity of the next hop candidate and dVelocity is the destination velocity. The rational between the concepts of the angle direction is to maintain the connection between vehicles as long as possible by choosing the small value of all calculated φ_id.

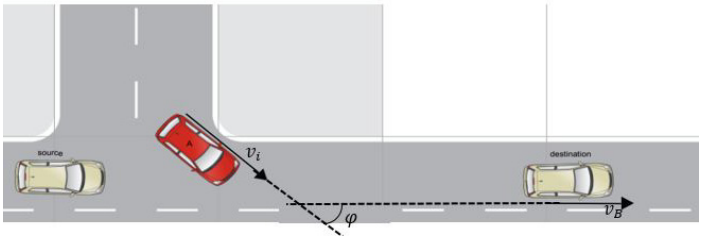


Figure 12 the angle direction φ

6. Conclusion

In this paper, we have provided a taxonomy of routing protocols in VANETS. Then we have given a quantitative analysis of the most known routing protocols for vehicular networks, using two simulation tools SUMO as a road traffic simulator and NS-3 as a network simulator, to determine the benefits and challenges of each algorithm.

From the results of the simulation, we have picked up the feebleness of GPSR and GPCR in terms of PDR and throughput. Based on the given results and to enhance the performance of those protocols, we have proposed a technique that combines the simple greedy forwarding technique and some other parameters selected carefully such as the direction, the speed and the density to choose the next relaying vehicle.

As a future work, we plan to develop our proposed technique to enhance GPSR and GPCR then prove their performance compared to the original GPSR and GPCR.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research is performed inside the MATSI Lab., ESTO, University Mohammed First, Oujda (Morocco).

References

[1] B. Amina and E. Mohamed, "Performance Evaluation of VANETs Routing Protocols Using SUMO and NS3," in *Colloquium in Information Science and Technology, CIST*, 2018, vol. 2018-Octob, pp. 525–530.

[2] A. Bengag and M. El Boukhari, "Classification and comparison of routing protocols in VANETs," in *2018 International Conference on Intelligent Systems and Computer Vision, ISCV 2018*, 2018, vol. 2018-May.

[3] E. Spaho, M. Ikeda, L. Barolli, and F. Xhafa, "Performance comparison of OLSR and AODV protocols in a VANET crossroad scenario," *Lect. Notes Electr. Eng.*, vol. 253 LNEE, pp. 37–45, 2013.

[4] S. Singh, P. Kumari, and S. Agrawal, "Comparative Analysis of Various Routing Protocols in VANET," *2015 Fifth Int. Conf. Adv. Comput. Commun. Technol.*, pp. 315–319, 2015.

[5] B. V. Coutinho, E. C. G. Wille, and H. I. Del Monego, "Performance of routing protocols for VANETs," *Proc. 9th Int. Conf. Ubiquitous Inf. Manag. Commun. - IMCOM '15*, pp. 1–6, 2015.

[6] R. Bala and C. R. Krishna, "Scenario Based Performance Analysis of AODV and GPSR Routing Protocols in a VANET," in *2015 IEEE International Conference on Computational Intelligence & Communication Technology*, 2015, pp. 432–437.

[7] A. Ashtaiwi and S. Abdusadik, "Performance Comparison of Position based Routing Protocols for VANETs," pp. 78–84, 2017.

[8] G. Sallam and A. Mahmoud, "Performance Evaluation of OLSR and AODV in VANET Cloud Computing Using Fading Model with SUMO and NS3," *2015 Int. Conf. Cloud Comput. ICCC 2015*, pp. 1–5, 2015.

[9] Z. J. Haas, M. R. Pearlman, and P. Samar, "The Zone Routing Protocol (ZRP) for Ad Hoc Networks," *draftietfmanetzonezrp02 txt*. p. 11, 2002.

[10] J. Liu, J. Wan, Q. Wang, P. Deng, K. Zhou, and Y. Qiao, "A survey on position-based routing for vehicular ad hoc networks," *Telecommun. Syst.*, vol. 62, no. 1, pp. 15–30, 2016.

[11] Jing Zhao and Guohong Cao, "VADD: Vehicle-Assisted Data Delivery in Vehicular <emphasis emphasistype="italic">Ad Hoc</emphasis> Networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1910–1922, May 2008.

[12] P.-C. Cheng, K. C. Lee, M. Gerla, and J. Härri, "GeoDTN+Nav: Geographic DTN Routing with Navigator Prediction for Urban Vehicular Environments," *Mob. Networks Appl.*, vol. 15, no. 1, pp. 61–82, Feb. 2010.

[13] Y. Lin, Y. Chen, and S. Lee, "Routing Protocols in Vehicular Ad Hoc Networks A Survey and Future Perspectives.," *J. Inf. Sci. ...*, vol. 932, no. c, pp. 913–932, 2010.

[14] B. Marzak, H. Toumi, E. Benlahmar, and M. Talea, "Analysing Clustering Routing Protocols Performance for Vehicular Networks," *Proc. 2nd Int. Conf. Big Data, Cloud Appl. - BDCA'17*, pp. 1–7, 2017.

[15] Y. L. Y. Luo, W. Z. W. Zhang, and Y. H. Y. Hu, "A New Cluster Based Routing Protocol for VANET," *Networks Secur. Wirel. Commun. Trust. Comput. NSWCTC 2010 Second Int. Conf.*, vol. 1, pp. 176–180, 2010.

[16] S. Allal, S. Allal, and S. Boudjit, "Geocast Routing Protocols for VANETs : Survey and Geometry-Driven Scheme Proposal Geocast Routing Protocols for VANETs : Survey and Geometry-Driven Scheme Proposal," no. February 2013, 2015.

[17] A. Bachir and A. Benslimane, "A multicast protocol in ad hoc networks inter-vehicle geocast," *IEEE Semiannu. Veh. Technol. Conf.*, vol. 4, pp. 2456–2460, 2003.

[18] Y. S. Chen and Y. W. Lin, "A mobicast routing protocol with carry-and-forward in vehicular ad hoc networks," *Int. J. Commun. Syst.*, vol. 27, no. 10, pp. 1416–1440, 2014.

[19] S. Zeadally, R. Hunt, Y.-S. Chen, A. Irwin, and A. Hassan, "Vehicular ad hoc networks (VANETS): status, results, and challenges," *Telecommun. Syst.*, vol. 50, no. 4, pp. 217–241, 2012.

[20] B. Williams and T. Camp, "Comparision of broadcasting techniques," pp. 194–205, 2002.

[21] B. Bako, "Efficient Information Dissemination in VANETs," 2016.

[22] O. Tonguz, N. Wisitpongphan, F. Bai, P. Mudalige, and V. Sadekar, "Broadcasting in VANET," in *2007 Mobile Networking for Vehicular Environments, MOVE*, 2007, pp. 7–12.

[23] M. Slavik and I. Mahgoub, "Spatial distribution and channel quality adaptive protocol for multihop wireless broadcast routing in VANET," *IEEE Trans. Mob. Comput.*, 2013.

[24] Q. Yang and L. Shen, "A Multi-Hop Broadcast scheme for propagation of emergency messages in VANET," in *International Conference on Communication Technology Proceedings, ICCT*, 2010.

[25] B. Bako, E. Schoch, F. Kargl, and M. Weber, "Optimized position based gossiping in VANETs," in *IEEE Vehicular Technology Conference*, 2008.

[26] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Proceedings - WMCSA'99: 2nd IEEE Workshop on Mobile Computing Systems and Applications*, 1999, pp. 90–100.

[27] T. Clausen and P. Jacquet, "Optimized Link State Routing Protocol (OLSR)," *Ietf Rfc3626*, p. 75, 2003.

[28] C. E. Perkins and P. Bhagwat, "Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 24, no. 4, pp. 234–244, 1994.

[29] G. He, "Destination-sequenced distance vector (DSDV) protocol," *Netw. Lab. Helsinki Univ. Technol.*, pp. 1–9, 2002.

[30] B. Karp and H. Kung, "GPSR: Greedy Perimeter Stateless Routing for wireless networks," *ACM MobiCom*, no. MobiCom, pp. 243–254, 2000.

[31] C. Lochert, M. Mauve, H. Füßler, and H. Hartenstein, "Geographic routing in city scenarios," *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 9, no. 1, p. 69, 2005.

[32] R. Manual *et al.*, "ns-3 Tutorial," *System*, no. January, pp. 1–46, 2011.

[33] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "SUMO – Simulation of Urban MObility," *Iaria*, no. c, pp. 55–60, 2011.

[34] OpenStreetMap Foundation, "OpenStreetMap," *Open Database License (ODbL)*. 2013.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 86-92 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Modelling the Methodology to Assess the Effectiveness of Distributed Information Systems

Ivan Aleksandrovich Tarkhanov[1,2,*], Galina Pavlovna Akimova[1], Matvey Aleksandrovich Pashkin[1], Aleksander Vladimirovich Soloviev[1]

[1]*Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, 117317, Moscow, Russia*

[2]*State Academic University for Humanities, 119049, Moscow, Russia*

ARTICLE INFO

Article history:
Received: 21 September, 2019
Accepted: 23 December, 2019
Online: 15 January, 2020

Keywords:
Effectiveness
Reliability
Information system
Geographical distribution
Reliability scheme
Prim algorithm

ABSTRACT

This research aims to develop the methodology to assess effectiveness of distributed information systems (hereinafter IS). The effectiveness of information systems is reviewed along with their reliability. The article underlines the importance of reliability assessment for large geographically distributed systems. We suggest the approach based on the hierarchical presentation of reliability scheme that constitutes the core of the mathematical model for reliability of a distributed system. The key advantage of the suggested approach is its applicability on geographically distributed systems of different architecture. Successful testing during effectiveness and reliability assessment on 3 geographically distributed IS allows to suggest that the developed methodology is ready to be used in IS reliability assessment projects.

1. Introduction

Rapid development of information and communication technologies contributed to the design and development of new large-scale geographically distributed systems that have become an integral part in the progress of state and transnational corporations. A large-scale geographically distributed system can be viewed as a set of interconnected hard and software. If an IS is unreliable or ineffective then a failure of one or several of the critical nodes can crash the whole system. Thus, it is crucial to be able to detect risks in due time and check operation performance of an IS. Regular reliability and effectiveness assessment will help to eliminate these risks [1]. In large geographically distributed IS, these problems are aggravated by the complexity and heterogeneity of IS hardware and software architecture. Thus, effectiveness and reliability assessment becomes especially important [2].

However, if the term reliability is quite clearly defined [3] then the term IS effectiveness is rather vague. Thus, according to the Russian state standard - GOST 34.003–90 – the effectiveness of an automated system means «the extent to which the goal set at the development of an IS is reached» [4]. In order to be able to measure IS effectiveness at the start of IS design developers have to define quality parameters by which system effectiveness will be assessed during operation [5], [6].

The following parameters are specified when an IS operational effectiveness is discussed from a technical point of view [7]:

- Adaptability – the IS property defining the ability to correlate to different functionalities and change accordingly;
- Availability - the IS property defining the availability of system functions and data stored in it at any time;
- Response time - the IS property defining the ability to guarantee timeliness and appropriate performance with consideration of declared performance indicators [8];
- IS reliability;
- Some studies identify other parameters, e.g. usability [9] or information security [10], resistance to external actions, including disaster recovery etc.

Practice has proved that the system of the abovementioned parameters is quite universal [7]. Besides, if we set one of the parameters as a primary one, e.g. reliability, then the rest of the parameters will be considered as restrictions. Alternatively, reliability may be considered as a restriction to other quality parameters. For example, the system may be "absolutely reliable"

* Ivan Aleksandrovich Tarkhanov, Email: itarkhanov@gaugn.ru

but it does not ensure availability of data or timeliness (reasonable response time), thus it cannot be considered effective.

Over the last 25 years there have been many researches dedicated to reliability in the context of IS architecture, data inside this IS and automated processes [5], [6]. However, none of these researches allows to clearly correlate the system reliability with certain distributed nodes and get a unified assessment that can be used in different projects. This article continues the reliability assessment research [11] and is dedicated to the modelling of methodology to assess effectiveness of geographically distributed information systems. The main quality parameter for the purposes of this article is IS reliability, while other parameters are regarded as restrictions.

2. Modelling the reliability of geographically distributed information systems

Reliability is definitely an important parameter of IS effectiveness. According to [12], reliability is an IS property enabling to preserve in time assigned parameters required to fulfill the main function under the influence of malfunctions (failures and breakdowns) of hardware, software and data errors, personnel and users mistakes in standard operational conditions and working environment when maintenance and servicing parameters are specified.

IS consisting of many parts distributed over a large territory have much longer operational lifespan than general life cycle for computer technologies that is normally 5-7 years [13, 14]. For example, reviewed herein IS - Russian Federation Automated System "Vybory" and "Electronic archive for personal record keeping" of the Russian State Pension Fund –have been operating not less than 20 years [15].

2.1. Task setting for modelling of the methodology to assess effectiveness of distributed information systems

IS reliability assessment in a dynamic environment with the purpose of detection of vulnerable spots in terms of reliability will allow to identify beforehand the areas of possible failures. Besides, it will enable to make a targeted modernization of an IS, and, finally, to save money. Providing this, the task setting for IS reliability assessment method can be formulated as follows:

Given:

1. The set of soft and hardware of an IS (TSSH) $PT = \{ PT_i \}$
2. Parametric perturbations on the PT set in time t:
3. $\varepsilon(t)$ – reliability violation on the PT set,
4. Requirements towards reliability parameters $T = \{ \rho T \}$ on the PT set

Find:

- A set of reliability parameters assessment models $PT : \rho$

Solution:

- Make an IS reliability scheme.
- Develop a general reliability model for this scheme.
- Develop a reliability assessment method.

Supplementary notes:

Reliability can serve as a restriction for other IS quality parameters: availability, response time etc.

Within this task, we do not consider the issues of mass nodes failures of a distributed IS for reasons of catastrophic character.

2.2. The modelling of a reliability scheme for geographically distributed information systems

The development of reliability scheme is the key stage at effectiveness modelling in general and reliability modelling specifically. It is convenient to introduce a reliability scheme for hierarchical geographically distributed IS in the form of graph $G(V, E)$, where $V = \{ v_i \}$ is a set of reliability nodes v_i. Every node consists of all soft and hardware stored on a certain closed territory (an office, a building, an organization). $E = \{ e_j \}$ is a set of communication channels (CC) e_j. Every element of the set E connects a couple of elements of the set V.

The development of a mathematical reliability assessment model by (network) graph is a difficult task. However, the majority of geographically distributed IS is designed hierarchically, which makes the task considerably easier.

Statement. Reliability scheme of a large hierarchical geographically distributed IS can be introduced in the form of a tree or a forest.

Let us consider an IS structure consisting of one central node and two regional nodes. Let us assume that CC e_j have communication bandwidth of 4 Mb/s on the region- center level, 2 Mb/s between regions. The formed reliability graph is a network consisting of three nodes and three arcs (see fig.1). Let us assume that data processing power on v_i nodes of the center-region graph correlates as 2:1 (proportionally to the number of lower-level nodes to one upper-level node). It is useless to make evenly powerful nodes because this will result in overspending at the development of an IS.

The use of Prim algorithm [15] to form a graph tree (in our case a maximum tree) with the starting vertex in the root of the tree (vertex Center) produces a spanning tree without region-region arc. Maximum tree is formed because there is a need to create communication channels that will guarantee implementation of timeliness restrictions at reliability assessment.

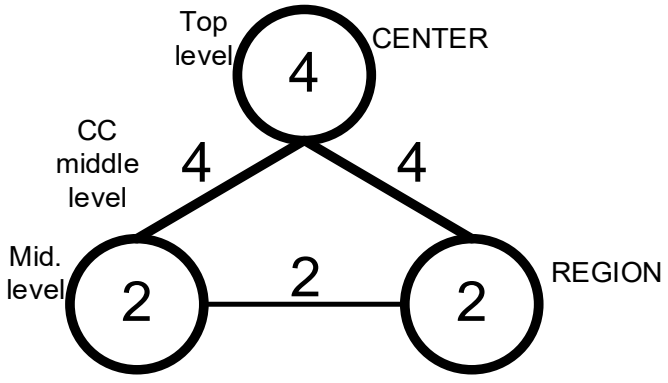


Figure 1. Reliability graph for basic two-level IS scheme

Increasing the number of levels in the IS hierarchy by one will produce an IS containing 3 levels of reliability nodes hierarchy v_i: center, region, district. In this IS every region has two districts. Let us assume that CC e_j have communication bandwidth of 4 Mb/s on the region-center level, 2 Mb/s on the district – region level, 1 Mb/s between districts of the same region, and 2 Mb/s between regions.

The formed reliability graph is a network consisting of seven nodes and nine arcs (see fig.2). Let us assume that data processing power on v_i nodes of the center-region-district graph correlates as

4:2:1 (proportionally to the number of lower-level nodes to one upper-level node).

Using the same modified Prim algorithm [16] we produce a spanning graph tree without arcs region-region and district-district. Thus, we got the same result for a three-level scheme as for a two-level one (providing we take into account the restriction that CC between nodes of the same hierarchy level are less powerful than CC between nodes of different hierarchy levels).

By applying a mathematical induction method, i.e. increasing the number of levels by addition of child nodes to the leaf node (fig.2) it is possible to form more complex graphs and maximum trees for them.

Providing the above we can formulate the task for reliability scheme modelling in the form of a tree or a forest based on the IS graph scheme.

Given:

A graph describing the IS scheme $G(V, E)$. Where $V = \{ v_i \}$ is a set of IS reliability nodes . Every node $v_i = \{ TSSH_{ik} \}$, $TSSH_{ik}$ is a certain set of soft and hardware of a node. $E = \{ e_j \}$ is a weighted set of all communication channels (CC) e_j. Every element of the set E connects a couple of elements of the set V. The weight of e_j means CC communication bandwidth.

Find:

The set of tree-type reliability schemes $Tr = \{ G'_l(V, E') \}$, where $G'_l(V, E')$ is the maximum tree l, $V = \{ v_i \}$ is the set of all reliability nodes v_i. $E' = \{ e_j \}$ is the set of IS communication channels (CC) of the maximum total weight that does not form a cycle. Every element of the set E' connects a couple of elements of the set V.

Restrictions:

- IS is hierarchically designed;
- The tree is formed by passing the arcs $G(V, E)$ from the root to the leaves (downwards) with the ban to pass the arcs upwards along the tree.

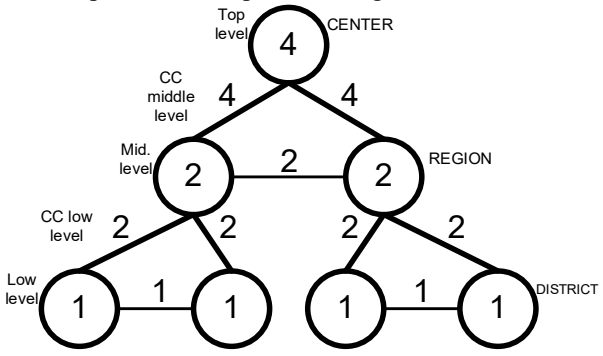


Figure 2. Reliability graph for a basic three-level IS scheme.

When a graph has a complex structure (see fig. 3), it may be difficult to locate a maximum tree (trees) because the use of classic Prim algorithm allows inclusion into the tree the arcs from lower levels to upper levels. This is possible, for example, when a starting vertex is Center 3 and there is no CC 5 (see fig.3). It is also possible to get a tree without arcs from vertex Center 3 to regions (see fig.3). Such solution will contradict the flow of main data in an IS.

For such cases, the Prim algorithm should be modified either by addition of the option allowing marking vertices in the data structure as non-visitable (e.g. in case of a failure) or the introduction into the algorithm of a vertex level label and a forced ban to consider arcs from lower to upper levels of the hierarchy (see fig.4).

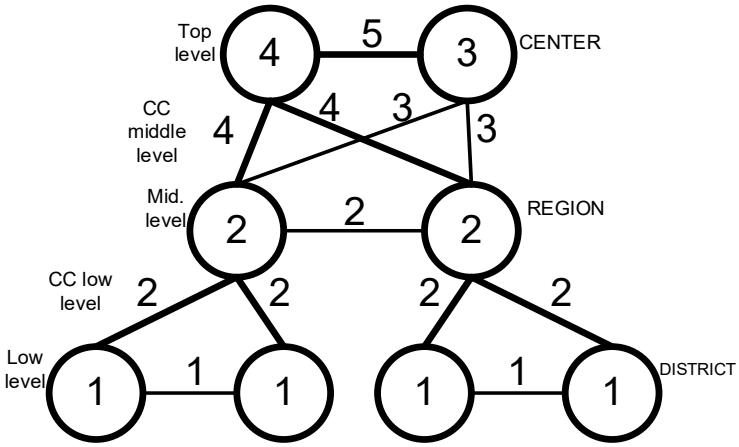


Figure 3. Reliability graph for a three-level IS scheme with 2 centers

If there is a backup data processing center for the central node (e.g. nodes Center 4 and 3 on fig. 3) then end-user processes are still addressing the same data copy at a time. Thus, the central node is considered as single (e.g. by combining vertexes Center 4 and 3 on fig. 3) but in the model it is necessary to consider its backup (e.g. in the basic case using the formula $K_{r.b.} = 1 - (1 - K_r)^n$, where n is the number of backup nodes for which reliability is defined as K_r).

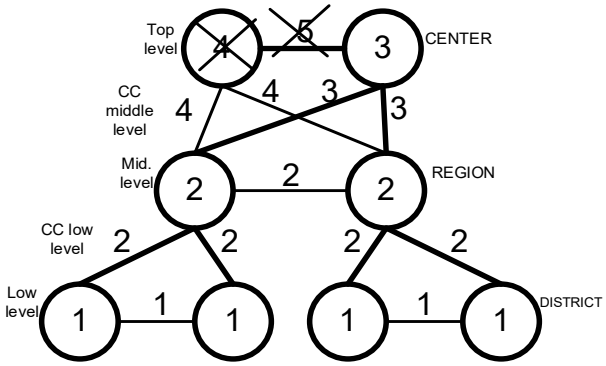


Figure 4. Reliability graph for a three-level IS scheme with two centers and the option to eliminate some vertices and arcs

If central nodes and communication channels have nominally the same processing power, it is possible to form several different trees if we choose different starting graph vertices, i.e. it is possible to form not a tree but a forest. Forming a forest does not significantly complicate the initial task. According to the reliability assessment model presented below it is necessary to make calculations by several trees and take the minimal assessment.

Forming one or several maximum trees (critical sections) in the general reliability graph is justified because the calculation and reliability scheme is much easier and there is no significant distortion of reliability assessment parameters.

Generally, for every tree (or section) of the initial reliability scheme there is a need to assess reliability according to the further suggested reliability assessment model and choose a minimal model. The complexity of reliability scheme algorithmic modelling depends on the number of vertices in every maximum

tree and in general case is $O(mn^2)$, where n is the number of vertices in a tree, and m is a number of trees.

The authors suggested this approach after the analysis of application of reliability assessment methods for large geographically distributed IS [17].

2.3. General reliability assessment model

IS reliability model for v_i nodes in general is formed based on the reliability scheme. Then, it is detailed for every v_i node depending on available soft and hardware $PT_i = \{ PT_{ij} \}$ of the v_i node.

The general reliability model scheme for a three-level hierarchical distributed IS (see fig.2) is as follows [18]:

$$\rho = K_{r\,IS} = K_{r\,t.l.} \sum_{i=1}^{Nm.l.} (b_i K_{r\,cc\,m.l.\,i} K_{r\,m.l.\,i} (\sum_{j=1}^{Nl.l.i} a_{ij} K_{r\,cc\,l.l.\,ij} K_{r\,l.l.\,ij}))$$

where $K_{r\,t.l.}$ – a reliability parameter for IS top-level objects (a non-detailed object reliability model for objects like data processing centers, central servers, software of central database etc.);

$b_i = Ob_i / \sum_{i=1}^{Nm.l.} Ob_i$ – the share of objects serviced by i- node of the middle level IS, Ob_i – a number of reliability elements serviced by i-node of the middle level IS ($\sum_{i=1}^{Nm.l.} b_i = 1$);

$N_{m.l.}$ – the number of elements of the middle level IS;

$K_{r\,cc\,m.l.\,i}$ – TSSH CC reliability parameter between top and middle level;

$K_{r\,m.l.\,i}$ –parameter of i-node reliability of the middle level IS;

$a_{ij} = Ob_{ij} / \sum_{j=1}^{Nl.l.i} Ob_{ij}$ – the share of objects serviced by j-node of the low level IS node connected with the middle level i ($\sum_{j=1}^{N\,l.l.i} a_{ij} = 1$) against general number of low level objects connected with i-node of the middle level IS;

$N_{l.l.i}$ – the number of low-level nodes connected with i-node of the middle level;

$K_{r\,cc\,l.l.\,ij}$ - the TSSH CC reliability parameter of the middle level i-node – low level j-node;

$K_{r\,l.l.\,ij}$ – the reliability parameter of the low level j-node connected with the middle level i-node.

A general model for reliability assessment is formed for availability parameter K_r (probability that the IS will be functioning at any given time), i.e. for a complex parameter for failure-free performance and recoverability of an IS. However, the choice of a reliability parameter is not crucial in this situation – the methodology shows how to design a reliability scheme, how to transform it into a tree (or a forest), how to form a reliability assessment model according to the scheme (schemes). Similarly, it is possible to form a reliability models for other parameters, such as:

- Operational availability parameter (K_{or}) is the probability that the object will be functioning at any given time t, apart from planned periods when the use of the object for its intended purpose is not required. Starting that moment the object will demonstrate failure-free performance for the set period.
- Technical use parameter (K_{tu}) is the ratio of expected value of intervals when the object is operationally available over a period of use to the sum of expected values of intervals when the object is available and idle due to technical maintenance or repairs over the same period of use.
- Planned use parameter (K_n) is a part of use period when the object must not be on planned maintenance or repair.
- Effectiveness parameter (K_{ef}) is a ratio of the value of effectiveness parameter when the object is used for its intended purposes for a certain period of operation to the nominal value of this parameter calculated under the condition that the object failures does not happen over the same period. The effectiveness parameter specifies the influence of object failures on its effectiveness when used for its intended purposes.

The suggested model for reliability assessment allows by induction to increase or decrease the number of levels, by calculation of corresponding weighting factors and by addition of the next level standard reliability parameters sum. It should be noted that the failure of a node of any level equals to the failure of all its child nodes. This partly takes into account mass failures.

3. Reliability assessment methodology

Now that the mathematical reliability assessment model is developed, it is time to start modelling the methodology to ensure reliability. It should be started with the elaboration of the terms failure and crash for a certain system, the choice of reliability parameters (availability rate, probability of failure-free performance, operational availability rate).

For the purpose of this research, IS is considered as a complex geographically distributed multifunctional system. Thus, in addition to «Full availability» and «Total crash» states there are many intermediate states with different levels of availability and corresponding performance effectiveness.

The developed model allows defining an IS total crash state - a total system crash is possible at the failure of upper level object and/or failure of all middle level objects and/or all low level objects.

Partial IS failure – the failure of one middle level object, which equals to the failure of all lower level objects handled by this middle level sector of an IS.

3.1. Modelling the methodology for reliability assessment

Summing up the methodology modelling in three major projects (Russian State Automated System “Vybory” [15], Electronic archive for personal record keeping” of the Russian Pension Fund and management system for the concept «Smart City Skolkovo» [16]) we came to the conclusion that it is necessary to undertake the following steps to develop a quantitative reliability assessment methodology and make the actual assessment:

1) Develop a reliability scheme in the form of a graph and detect one or several trees.
2) For every formed tree develop a mathematical model for reliability assessment according to the presented above model specifying the number of node levels.
3) Elaborate the terms for failure and crash of an IS according to its functionality. For example, the system failure is registered if function $f_i \in F$ cannot be executed for a period of time more than T minutes (seconds), where F is a set of functions, crucial to IS if not executed.
4) Define the terms failure and crash according to the critical number of IS operating nodes. For example, the failure of an IS node is registered if less than N automated working

stations (AWS) of an IS does not perform required functionality F for a period of time more than T minutes (seconds).

5) Define reliability parameter (parameters), principles of failure and crash probability distributions, restrictions and assumptions at reliability assessment.
6) Define the effect other IS have on the system reliability and adjust the mathematical model taking into account this effect.
7) Develop the form and procedure for the collection of IS reliability statistics data, which include every node and element of IS scheme and reliability model.
8) Collect statistic data on the nodes and reliability elements of the IS.
9) It is recommended to develop mathematical models for reliability parameters, e.g. for timeliness and IS function execution, and the scheme for the use of these parameters at the system reliability assessment. For example, if at the current reliability rate timeliness is not sufficient, then the failure of a node or IS in general is registered. Thus, in the most general case reliability parameters serve as reliability model restrictions.
10) Specify the general mathematical reliability model by every node and reliability element taking into account possibility of backup and influence of other parameters and systems.
11) Calculate system reliability parameters taking into account restrictions (including timeliness and function execution). The calculations are made for every formed i-tree. Then a pessimistic approach is applied, i.e. $K_r\ IS = min(K_r\ IS_i)$ is chosen.
12) Calculate reliability parameters according to the model described above for every node of every level of the reliability scheme.
13) For reliability nodes for which there are no failure statistics calculate anticipated reliability as the lowest among the same calculated reliability nodes of an IS.
14) If necessary, specify parameters calculations for every node or reliability element.
15) At the end of the calculation make an IS "health chart", i.e. IS scheme where risky nodes and reliability elements are marked.
16) The overall risk and weaknesses chart used at the detailed reliability assessment is defined by detected critical nodes.
17) If the risks are confirmed by the detailed assessment, it is necessary to develop an action plan to eliminate these risks.

3.2. Specifics of the interpretation of calculation results in the reliability model

Consider a more complex case. A distributed IS consists of several top level nodes, thus it is necessary to form a maximum tree for every vertex and calculate the assessment for every such tree.

As a result we have a set of assessments: $K_{r\ IS} = \{ K_{r\ ISi} \}$.

In IS reliability assessment we apply a pessimistic approach, i.e. choose $K_{r\ IS} = \min_i(K_{r\ ISi})$.

The produced indicators both for IS in general and for every separate reliability node and element are compared against the reliability set $T = \{ \rho_T \}$ existing on the PT set.

Reliability assessment can be done quite fast with the use of the suggested method. And with the use of some software for calculations even faster. However, the assessment is not too accurate and can be used only for estimated rough calculations.

4. Modelling the effectiveness of distributed information systems

In order to assess effectiveness it is needed not only to model reliability but also develop models of other IS quality performance parameters.

The authors developed the following mathematical models of IS quality parameters for the assessment of IS effectiveness: IS availability and response time (timeliness). These models can be supplemented by IS reliability assessment models as additional restrictions.

4.1. The model for the assessment of IS functions performance availability

Here is the formal task formulation for the IS availability assessment by listed functions.

Given:

1) Functionality volume F (total number of all IS operations by all operation types).
2) Data volume D (total data volume stored in IS database).
3) Restriction F_d (maximum admissible F value that allows to consider IS functional).
4) Restriction D_d (maximum admissible D value that allows to consider IS functional).
5) IS reliability (K_r).

Find:

Probability of availability of the minimum admissible IS functionality and data volume.

Solution:

Mathematical model of availability is the probability to perform IS operation s by a user:

$$P_p\,(F \geq F_d) = P(F \geq F_d)\ K_{r\ IS}$$

And the probability of IS data availability to a user:

$$P_d\,(D \geq D_d) = P(D \geq D_d)\ K_{r\ IS}$$

where $P(F \geq F_d)$ – the probability of performing functionality volume F, i.e. performing of IS functions not less than directly set (or admissible) F_d subject to failure-free performance of IS soft and hardware;

$P(D \geq D_d)$ – the probability of availability of data volume D not less than directly set (or admissible) D_d subject to failure-free performance of IS soft and hardware;

$K_{r\ IS}$ – IS reliability indicator calculated according to the reliability assessment model (see above the general IS reliability assessment) that serves as a restriction of the model.

$$P(F \geq F_d) = n_i / N(i),$$

where n_i – the number of operation of i-type (see the list of operation types above) that satisfies the inequation $F \geq F_d$,

$N(i)$ – the total number of operation of *i-type* performed in an IS at the collection and processing of information.

$$P(D \geq D_d) = d / D,$$

where d – the volume of available data,

D – total data volume in an IS.

Additional parameters specifying availability:

- Average percent of IS functions F availability over a period of T;
- Average percent of IS data D availability over a period T.

4.2. The model for the assessment of IS functions timeliness (response time)

Here is the formal task formulation for the IS timeliness assessment.

Given:

1) Functionality volume F (total number of all IS operations by all operation types).
2) Restriction T_d (maximum admissible time value that provides an IS user with acceptable result of the IS function performance)
3) IS reliability ($K_{r\,IS}$)

Find:

The probability of IS performance of its functionality in full over a period not less than maximum admissible time

Solution:

Mathematical model for timeliness is defined as the probability of information processing over a period of t that is not more than the set (admissible) T_d, with reliability restriction:

$$P_t\ (t \le T_d) = P\ (t \le T_d)\ K_{r\,IS}\ ,$$

where $P\ (t \le T_d)$ – the probability of information processing during operations performance (the types of operations are listed in the first paragraph of the section) in an IS over a period of t that is mot more than the set (admissible) T_d subject to failure-free performance of soft and hardware;

$K_{r\,IS}$ – IS availability indicator (see section 2.4).

In order to calculate all time intervals T it is necessary to obtain statistics data and find admissible time T_d allowing to get the probability assessment of function timeliness.

Probability assessment of i-type function timeliness over a period not exceeding admissible is calculated by the following formula:

$$P(T_i \le T_{di}) = n_i / N(i)$$

where n_i – the number of performed i-type operations T_i that satisfy the inequation $T_i \le T_{di}$;

$N(i)$ – the number of operations of i-type from the set of operations $F\,l=[1, N(i)]$.

5. Practical implementation of reliability assessment methodology

Effectiveness and reliability assessment methods developed according to the suggested by the authors methodology are described in [13, 18]. This methodology can be easily automated. For example, software complex for the assessment of effectiveness and reliability of the Russian Federation Automated System "Vybory" was created based on this methodology.

The suggested methodology formed the basis for the methods used by the Russian Federation Automated System "Vybory" and soft and hardware complex "Electronic archive for personal record keeping" of the Russian Federation Pension Fund. The calculations made according to the implemented methods revealed reliability bottlenecks in IS nodes, which allowed to make their targeted modernization.

Besides, the developed methodology was used at the development of the logic-mathematical model for Skolkovo "Smart city" management system concept for the Skolkovo Innovation center [17]. This allowed to design the management system with regard to reliability and effectiveness on the concept stage and set justified criteria for the selection of soft and hardware solutions at the detailed design stage.

6. Comparison with other approaches and model restrictions

First researches on the reliability assessment of distributed systems were focused on the solution of individual tasks - assessment of mistakes at data transfer or instability of communication channels bandwidth between IS nodes [19],[20]. Engineering approach and special set of tools to assess IS reliability is described in [5] и [6].

The most detailed IS effectiveness assessment model that takes into account not only the quality of the system itself but also the data circulating in it, user experience and other business-oriented parameters is D&M IS Success Model [7]. According to classification [7] the described model assess only technical aspect of the system performance and is similar to approaches based on IS architecture assessment [20], [21]. It can be used along with D&M IS Success Model and supplement it with regard to geographically distributed systems.

Another possible application of the suggested method is its use with IS effectiveness assessment models based on machine learning and data analyses [22]. During machine learning the reliability scheme in force (or any other parameter of IS effectiveness) is adjusted and used for the presentation of the current situation and fast calculation of IS effectiveness parameters.

The majority of researches focuses on "reliability of data" circulating in the system (a process approach) [23] or on failure probability [24] not taking into account architecture specifics of the assessed IS.

The most recent researches on the reliability assessment of distributed systems are mainly based on the use of Markov chains [2], [25] that allow to understand the correlation between the states of different system components, define bottlenecks of a small distributed system. However, it does not allow to get a clear IS reliability assessment comprising over 100 nodes in a dynamic state.

In section 1.2 it is demonstrated that reliability scheme modelling with the use of Prim algorithm is applied for hierarchical systems. The same approach may be employed in the non-hierarchical systems. To do so it is needed to locate one or several central nodes and form the reliability scheme for each of them.

It should be noted that the suggested model does not take into account some important parameters, e.g. it does not cover the issue of data deterioration at the reliability assessment [26] that in future could be considered at the development of reliability scheme or any other parameter of IS effectiveness. A simplified description of the mathematical model was presented in the suggested methodology.

7. Conclusion

This article describes methodology for modelling IS effectiveness that was brought up to the level of applied operational methods that were implemented during the development of the major federal level information systems: the Russian Federation State Automated System "Vybory", "Electronic archive for personal record keeping" of the Russian Federation Pension Fund, Skolkovo "Smart city" concept, which is undoubtedly the advantage of this work.

Another advantage of the developed mathematical model is that it takes into consideration a geographically distributed architecture of an IS and has a great potential for extensibility into more complex distributed IS configuration.

In further studies, it is planned to improve Prim algorithm convergence at the formation of critical sections (maximum trees of the reliability scheme), to develop a method for specification of nodes reliability models of the reliability scheme and its automatization with the help of specialized software, to develop a methodology to model other IS quality parameters such as adaptability, resistance to external actions (including actions of catastrophic character).

We plan to develop the software allowing to improve the reliability scheme calculation of certain nodes and elaborate modelling methodology for such effectiveness parameters as adaptability and disaster recovery. The focus of the research will be on the analysis of Prim algorithm convergence at the calculation of several maximum trees.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] L. F. Pitt, R. T. Watson, C. B. Kavan "Service quality: a measure of information systems effectiveness" MIS quarterly, 173-187, 1995. https://doi.org/10.2307/249687

[2] A. Waseem, Y.W. Wu "A survey on reliability in distributed systems" Journal of Computer and System Sciences, No. 79.8., 1243-1255, 2013

[3] F. M. Zahedi "Reliability metric for information systems based on customer requirements" International Journal of Quality & Reliability Management 14.8, P. 791-813, 1997. https://doi.org/10.1108/02656719710181312

[4] GOST 34.003-90 Informacionnaja tehnologija. Kompleks standartov na avtomatizirovannye sistemy. Avtomatizirovannye sistemy. Terminy i opredelenija, M.: Izd-vo standartov., 1992.

[5] M. C. Boudreau, D. Gefen, D.W. Straub "Validation in information systems research: a state-of-the-art assessment" MIS quarterly, 1-16, 2001. https://doi.org/10.2307/3250956

[6] P.K. Kapoor, H. Pham, A. Gupta, P.C. Jha "Software reliability assessment with OR applications" Springer, London, 2011.

[7] W. H. Delone, E. R. McLean "The DeLone and McLean model of information systems success: a ten-year update" Journal of management information systems. No. 19(4). 9-30, 2003. https://doi.org/10.1080/07421222.2003.11045748

[8] X. Lu, K. Moriyama, I Luque., M. Kanda, Y. Jiang, R.Takanuki, K. Mori "Timeliness and reliability oriented autonomous network-based information services integration in multi-agent systems" IEICE TRANSACTIONS on Information and Systems. No. 88(9), 2089-2097, 2005. https://doi.org/10.1093/ietisy/e88-d.9.2089

[9] A. Bangor, P. T. Kortum, J. T. Miller "An empirical evaluation of the system usability scale" Intl. Journal of Human–Computer Interaction. No. 24 (6), 574-594, 2008. https://doi.org/10.1080/10447310802205776

[10] M. Nancy, T. Stehney "Security quality requirements engineering (SQUARE) methodology" ACM SIGSOFT Software Engineering Notes. Vol. 30. No. 4. 2005. https://doi.org/10.1145/1082983.1083214

[11] Akimova G. P., Solovyev A. V., Tarkhanov I. A., Reliability assessment method for geographically distributed information systems // 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), Almaty, Kazakhstan, 2018, pp. 1-4., https://doi.org/10.1109/ICAICT.2018.8747055

[12] GOST 27.002-89. "Reliability in technology. Basic concepts. Terms and Definitions", M.: Publishing Standards, 1989.

[13] P. Stafford "You can believe the Hype Cycle's take on technology" Financial Times. 30 May 2008.

[14] B. Furneaux, M. R. Wade "An exploration of organizational level information systems discontinuance intentions", Mis Quarterly, 573-598, 2011. https://doi.org/10.2307/23042797

[15] G. P. Akimova, E.V. Pashkina, A. V. Soloviev "Analiz ocenki jeffektivnosti ierarhicheskoj territorial'no-raspredelennoj informacionnoj sistemy na primere GAS "Vybory" Proceedings of the Institute of System Analysis of the Russian Academy of Sciences, vol. 58, 25-38, 2010.

[16] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein "Introduction to Algorithms" Third Edition. MIT Press. ISBN 0-262-03384-4. Section 23.2: The algorithms of Kruskal and Prim, 631–663, 2009.

[17] A.V. Soloviev, Concept «Umnyj gorod Skolkovo» for Skolkovo innovation center and set of measures for its implementation. Vol 3. Logiko-matematicheskaja model' «Umnogo goroda»: research report. ISA RAN. 2012.

[18] A. V. Soloviev, E.Y. Komleva, "Metodicheskoe obespechenie nadezhnosti v oblasti hranenija jelektronnyh dokumentov" Proceedings of the XXIII International Conference "Documentation in the Information Society: Archival Studies and Documentation in the Modern World", Rosarchive, , 321-331, 2017.

[19] F. Proschan, R. E. Barlow "Mathematical theory of reliability." SIAM series in applied mathematics. 1967.

[20] L. Svobodova "Reliability issues in distributed information processing systems" Dig. Papers FTCS-9: 9th Annu. Symp. Fault-Tolerant Computing. 1979.

[21] K. Goševa-Popstojanova, S. T. Kishor "Architecture-based approach to reliability assessment of software systems" Performance Evaluation. No. 45.2-3. 179-204. 2001. https://doi.org/10.1016/s0166-5316(01)00034-7

[22] N. P. L. Ernawatiningsih, P. Kepramareni,, "Effectiveness of Accounting Information Systems and the Affecting Factors" International Journal of Applied Business and International Management, 4(2), 33-40, 2019 https://doi.org/10.32535/ijabim.v4i2.564

[23] R. Krishnan, J. Peters, R. Padman, D. Kaplan "On data reliability assessment in accounting information systems" Information Systems Research. No. 16(3). 307-326, 2005. https://doi.org/10.1287/isre.1050.0063

[24] A. Choudhary, A.S Baghel., O.P. Sangwan "Parameter Estimation of Software Reliability Model Using Firefly Optimization" Data Engineering and Intelligent Computing. Springer. Singapore, 407-415, 2018. https://doi.org/10.1007/978-981-10-3223-3_39

[25] Z. Zheng "Semi-markov models of composite web services for their performance, reliability and bottlenecks" IEEE Transactions on Services Computing, No. 10.3, 448-460, 2017. https://doi.org/10.1109/tsc.2015.2475957

[26] J. Xu, D. Yu, Q. Hu, M. Xie "A reliability assessment approach for systems with heterogeneous component information" Quality Engineering, 1-11, 18.12.2017. https://doi.org/10.1080/08982112.2017.1402935

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 93-99 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Constrained Path Planning for Both Smooth Motion Profile and Stable Control

Anh Son Tran[1], Ha Quang Thinh Ngo[2,*]

[1]*Dept. of Manufacturing Engineering, HCMC University of Technology, Vietnam National University (VNU-HCM), Vietnam*

[2]*Dept. of Mechatronics Engineering, HCMC University of Technology, Vietnam National University (VNU-HCM), Vietnam*

ARTICLE INFO

Article history:
Received: 23 October, 2019
Accepted: 10 December, 2019
Online: 15 January, 2020

Keywords:
Polynomial Interpolation
Automation
Vibration-less

ABSTRACT

In our era, human and machine must share the same workspace in various plants such as warehouse logistics, material transportation, manufacturing factory or assisted human. The collaboration is essential to maintain the safe operation and the complete mission. In this paper, a design of anti-vibration scheme using Spline generator for grounded vehicle platform in human-machine interaction is introduced. This hardware structure is able to play a role as carrier in loading application or human assistant in real life. In most of cases, vehicle must convey cargo from start point to destination instead of human while its reference trajectory conflicts with human's movement. The proposed algorithm helps to carry out a decision of stopping strategy with the least vibration as possible to avoid accident. From the test results, the proposed design is feasible and capable to apply in numerous industries.

1. Introduction

In the first and second industrial revolution, human labor is the main factor to build the products in the workshops, factories, or warehouses. The productivity of this stage mostly depends on the human factor such as clock operation, time off, overtime, short-term payment. Until 1950s, the Automated Guided Vehicle (AGV) was firstly introduced as an advanced mobile robot which uses wires, UV marks and light to navigate. Nowadays, AGVs are even more advanced by using magnets, laser range finder or digital cameras to pilot. Moreover, the most complex guidance technique is currently an integrated on-board with sensors and software, using wireless or radio frequency connection for routing. This autonomous system is commonly applied in large scale factories and storehouses to transport materials from one place to another or assist human by following them with packages. AGVs come with powerful computers that are connected to the cameras, lasers and sensors. This allows vehicle to handle multi-tasking or complex missions all the time. It is completely safe to walk or stand around AGVs while they are operating. On the other hand, unlike human, AGVs can work overtime and will result in the growth of productivity and minimize the expenses.

However, in most of the autonomous system, while they are working, there are always exist residual vibration, which leads to the lack of accuracy. For instance, the crane with heavy load in the end-effector is moving from one place to another. And when it's stopped, the vibration occurs proportionally to the mass of the load, which not only lead to the inaccuracy but also damage the joints of the crane. In other example, the AGV is following human with a bulky material above and still maintain the safe distance to human. It is assumed that sudden collision would force AGV to stop immediately. If it is not stopped properly, the materials above would collapse due to the vibration. In [1, 2], the researchers composed of four major functional blocks including human detection and human feature, dynamic social zone, approaching pose of the robot to a human or a human group. They are incorporated into a motion planning system, comprising a local path planner and dynamic window technique. The output of research is only focus on driving a mobile robot to approach both stationary and moving humans in a socially acceptable manner. They did not concern on dynamic constraints of system model to guarantee the safe operation. In the applications where the oscillating dynamics is not easily detectable with the standard sensors that hung on the plant, accordingly, it cannot be compensated by a feedback control action. The use of a feed-forward control is often the only option for reducing the vibrations. this technique [3-5] is very effective in vibration suppression, but it may be not useful in trajectory following application such as 3D printing, cutting ... The alternative method to dominate oscillation

* Ha Quang Thinh Ngo, Email: nhqthinh@hcmut.edu.vn

is the use of finite impulse response filter for planning minimum-time trajectories [6, 7] was presented. This investigation connects between analytic functions and dynamic filters permit a generalization of various trajectories. It is usually obtained by second- or third-order polynomial functions to a generic order with only a modest increase of the complexity. The exponential jerk along with filters was investigated as a new type of trajectory [8, 9]. As a matter of fact, constant jerk trajectories did not guarantee a complete vibration suppression when the damping of the resonant modes is not negligible. The decay rate and duration of the jerk impulses that allow residual vibration cancellation are derived in an analytical way as a function of the dynamic characteristics of the plant. Some comparisons with the well-known input shaping techniques combining with filters performs the advantages of this method. In [10-12], the authors demonstrated the theories of motion profiles from basic to advance, the mechanism of motion generators and their applications in industry. They acquainted with polynomial profile, exponential profile and multi-segmented profile to figure out the generation of robot trajectory. One of advanced path generators is a uniform B-splines of degree p. The investigators in [13-15] developed the spline trajectories with a chained composed by p average filters to optimize. It could find out the solutions for two different troubles such that exactly interpolating a set of given points by means of a complex trajectory and the need of suppressing mechanical vibrations. However, these researches are only simulated in virtual environment from theories. It needs more time to overcome practical problems, for example, time consuming to generate motion profile, burden computation in computer or high cost comparing to other methods.

In this paper, the exponential B-spline trajectories are taken into consideration and able to generate exactly if the profiling parameters are chosen properly. Besides, for safety in the working environment, there are many options such as stereo cameras, laser sensors, proximity sensors that could be placed additionally. A Kinect camera is chosen because of its advantages, i.e. cheap price, easy to calibrate, library supported and most importantly human detectable. A Kinect camera is attached on the vehicle to calculate the distances when humans are detected while moving along its path, based on the depth sensor integrated in the camera. The experimental results performed on an automatic guided vehicle, show the effectiveness of the proposed method.

2. B-spline and Applications

Splines are piecewise polynomial functions widely used to interpolate sets of data points or to approximate functions, curves and surfaces. The data may be either one-dimensional or multi-dimensional. A spline of order p is a piecewise polynomial function of degree $p-1$ in a variable of t. The values of u where the pieces of polynomial meet are known as knots, denoted $t_0, \dots, t_{knot}$ and sorted into non-decreasing order. When the knots are distinct, the first $p-1$ derivatives of the polynomial pieces are continuous across each knot. When r knots are coincident, only $p-r$ derivatives of the spline are continuous across that knot.

2.1. B-spline Functions

A particularly efficient technique for the computation of splines is based on so called B-splines, or Basic-splines. The reason for such a name is that a generic spline can be obtained as a linear combination of a proper number of B-spline basis functions, $B_j^p(t)$, i.e.

$$s(t) = \sum_{j=0}^{m} p_j B_j^p(t) \tag{1}$$

where

$s(t)$ is the curve trajectory.

$p_j, j = 0, \dots, m$ are the control points.

$B_j^p(t)$ is the j^{th} B-spline basis function of degree p.

$t = [t_0, \dots, t_{knot}]$ is the knot vector.

B-spline basis function can be generated by using Cox-de Boor formula or convolution method. B-spline is a general term.

where

$B^{(0,0,0,0)}(t) = B^3(t)$ is a cubic polynomial spline.

$B^{(0,0,0,\widehat{\alpha})}(t)$ is an exponential spline of degree 3.

$B^{(0,0,\widehat{-\alpha},\widehat{\alpha})}(t)$ is an exponential tension spline of degree 3.

$B^{(\widehat{\alpha_0},\widehat{\alpha_1},\widehat{\alpha_2},\widehat{\alpha_3})}(t)$ is an exponential general spline form produces B-splines that are a linear combination of function $e^{\widehat{\alpha_0}t}, e^{\widehat{\alpha_1}t}, e^{\widehat{\alpha_2}t}, e^{\widehat{\alpha_3}t}$.

Cubic polynomial spline can be generated in both methods. Exponential spline can only be generated by using convolution method.

Cubic polynomial spline basis function using Cox-de Boor. In this method, B-splines are piecewise polynomial, characterized by a knot vector $t = [t_0, \dots, t_{knot}]$. The relation between knots, n_{knot} and number of control points, m is $n_{knot} = m + p + 1$.

$$B_j^0(t) = \begin{cases} 1 & t_j \le t < t_{j+1} \\ 0 & otherwise \end{cases} \tag{2}$$

$$B_j^p(t) = \frac{t - t_j}{t_{j+p} - t_j} B_j^{p-1}(t) + \frac{t_{j+p+1} - t}{t_{j+p+1} - t_{j+1}} B_{j+1}^{p-1}(t) \tag{3}$$

Exponential spline basis function using convolution. In this method, B-splines characterized by an equally-spaced distribution of the knots $t_{j+1} - t_j = T, j = 0, \dots, m-2$.

$$B^{\alpha_k}(t) = \begin{cases} \dfrac{\alpha_k T}{e^{\alpha_k T} - 1} e^{\alpha_k t} & 0 \le t < T \\ 0 & otherwise \end{cases} \tag{4}$$

$$B^{(\alpha_0,\alpha_1,\alpha_2,\alpha_3)}(t) = \frac{1}{T} B^{\alpha_3}(t) * \dots * \frac{1}{T} B^{\alpha_1}(t) * B^{\alpha_0}(t) \tag{5}$$

2.2. B-spline Applications

The linear combination of B-spline basis functions with a set of control points can solve the interpolation problem, therefore, in mathematical subfield of numerical analysis, it is widely used for curve fitting and numerical differentiation of experimental data. In computer-aided design and computer graphic, B-spline basis functions are used for generating and representing curves or surfaces.

3. Motion Planner Using B-Spline

Smoothness, flexibility and possibility of exactly interpolating some via-points are the main characteristics of B-spline functions. Therefore, we believe that it is possible to generate B-spline trajectories which exactly cross the desired via-points and at the same time are able to suppress vibrations.

3.1. B-spline Curve Generator

Using Cox-de Boor formula, it can only solve the cubic polynomial spline case. Though, it's a practical method to generate the B-spline trajectory for autonomous system while maintaining the properties of system. The proposed algorithm of flexible and controllable motion generator is illustrated in Fig. 1. In this case, the B-spline curve has degree 3 together with its control polygon defined by.

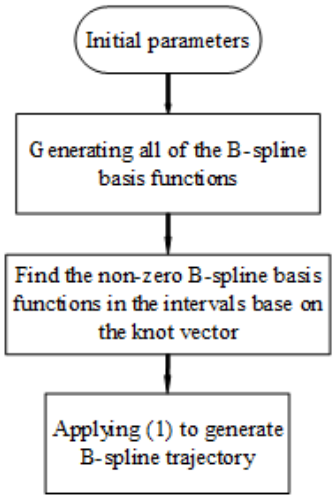


Figure 1. Proposed algorithm of B-spline trajectory generator.

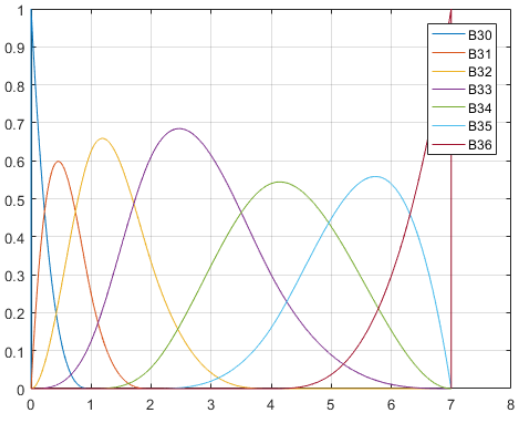


Figure 2. B-spline basis functions using Cox-de Boor formula

$$P = \begin{bmatrix} p_0, & p_1, & \cdots & p_{m-1}, & p_m \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & -3 & 4 & 5 & -5 & -6 \end{bmatrix} \quad (6)$$

The knot vector is,

$$u = \begin{bmatrix} 0, & 0, & 0, & 0, & 1, & 2, & 4, & 7, & 7, & 7, & 7 \end{bmatrix} \quad (7)$$

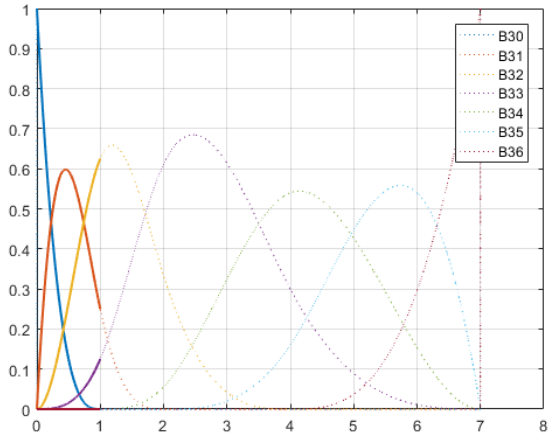


Figure 3. Non-zero B-spline basis functions in the interval [0,1].

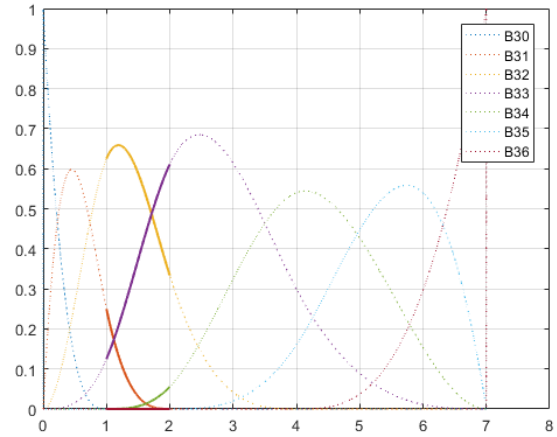


Figure 4. Non-zero B-spline basis functions in the interval [1,2].

To generate the B-spline with above parameters, a method of path planning using various basis functions has been investigated. According to Fig. 1, with the initial parameters (the control point vector and the knot vector) and a combination with (3), it is able to generate the B-spline basis functions, note that the variable t_j in (3) now becomes u_j in (7) with $j = 0, \ldots, 10$ with respect to the example. Since it has a total of seven control points, it requires seven B-spline basis functions of degree 3 which are $B_0^3(t), \ldots, B_6^3(t)$ to be able to calculate the trajectory using (1). After calculating all of the B-spline basis functions, note that in every knot span $[u_j, u_{j+1})$, there are at most $p + 1$ basis functions B_j^p are not null, namely $B_{j-p}^p, \ldots, B_j^p$, for example there are 4 basis functions $B_0^3, B_1^3, B_2^3, B_3^3$ are not null in the interval $[u_3, u_4)$ as shown in Fig. 5. Using (1), for each of the B-spline basis functions of degree 3 as mentioned above, multiply it with the corresponding control points, as a result we have $s(t)$, which is a vector with the first row stands for the time in the x-axis and second row stands for the amplitude of the trajectory in the y-axis.

Fig. 2 introduces all of the B-spline basis functions in the interval [0,7). Note that, at every sampling time, the sum of all of

the B-spline basis functions at that time is always equal to 1, this is also one of the characteristics of the B-spline basis functions.

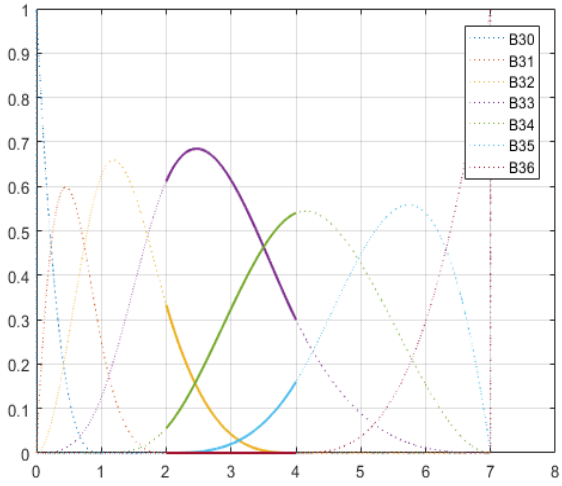


Figure 5. Non-zero B-spline basis functions in the interval [2,4].

In Fig. 3 and Fig. 4, there are 4 non-zero basis functions which are available in the interval $[0,1)$ and $[1,2)$, they are $B_0^3, B_1^3, B_2^3, B_3^3$ and $B_1^3, B_2^3, B_3^3, B_4^3$ respectively.

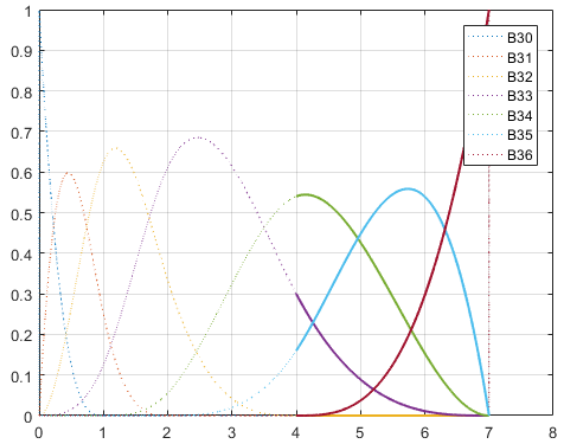


Figure 6. Non-zero B-spline basis functions in the interval [4,7].

3.2. Smooth Profiling Generation

In order to prove the effectiveness and correctness of B-spline application for vibration-less motion, a practical example of material transportation in factory has been chosen. An automated guided vehicle (AGV) with cargo must travel in long distance during working hours. It is recognized that the stable movement is difficult to gain since the vibration occurs. To conquer these troubles, we apply B-spline trajectory to reduce the oscillation. Fig. 10 illustrates the system modeling of both AGV and cargo as m_1 and m_2. They are linked together via a string (stiffness k) and damper (viscous coefficient b).

$\vec{F}_{qt}$: inertial force acting on m_2

$\vec{v}$: velocity of m_1

$\vec{x}_1, \vec{x}_2$: displacement of m_1 and m_2

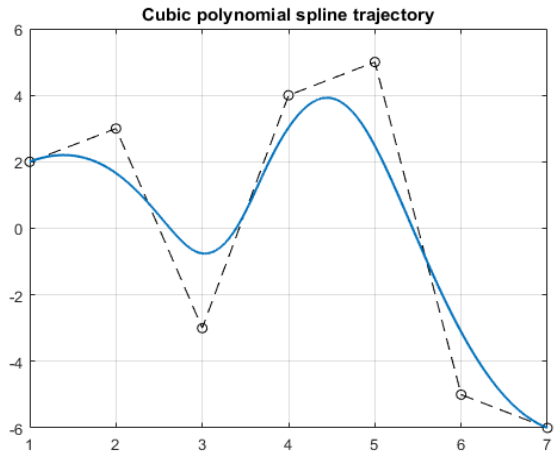


Figure 7. Cubic polynomial spline trajectory.

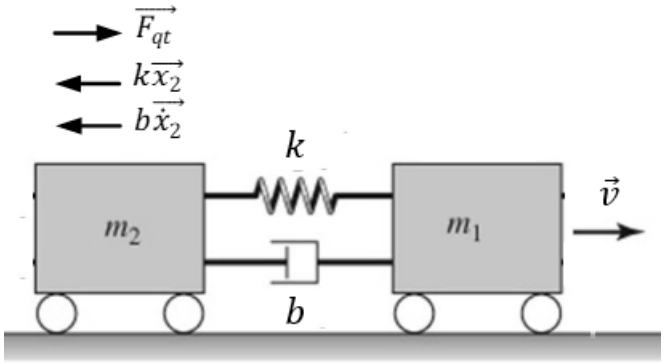


Figure 8. Modeling of both vehicle and cargo.

Using Newton law, we have equation of system dynamics in time domain

$$m_2\vec{a}_2 = \vec{F}_{qt} + k\vec{x}_2 + b\dot{\vec{x}}_2 \quad (8)$$

In its direction,

$$m_2\ddot{x}_2 = -kx_2 - b\dot{x}_2 + m_2\ddot{x}_1 \quad (9)$$

By applying Laplace transformation, the below description is in frequency domain

$$m_2 s^2 X_1 = m_2 s^2 X_2 + kX_2 + bsX_2 \quad (10)$$

$$\Leftrightarrow X_2 = \frac{s^2 X_1}{s^2 + 2\varepsilon\omega_n s + \omega_n^2} \quad (11)$$

$$\Leftrightarrow \frac{X_2}{X_1} = \frac{s^2}{s^2 + 2\varepsilon\omega_n s + \omega_n^2} \quad (12)$$

where,

$\omega_n = \sqrt{\frac{k}{m_2}}$: the natural frequency of the of the lightly damped system.

$\varepsilon = \frac{b}{2m_2\omega_n}$: the damping ratio.

From equation (11), it has been known that the system vibrates more at pole points while it maintains a stable state at zero points are the points. On the other hand, the system is stable if the poles lay on the left side of the complex plane or being replaced by the zeros using the pole placement technique. In (12), we can find the zeros and poles of the system.

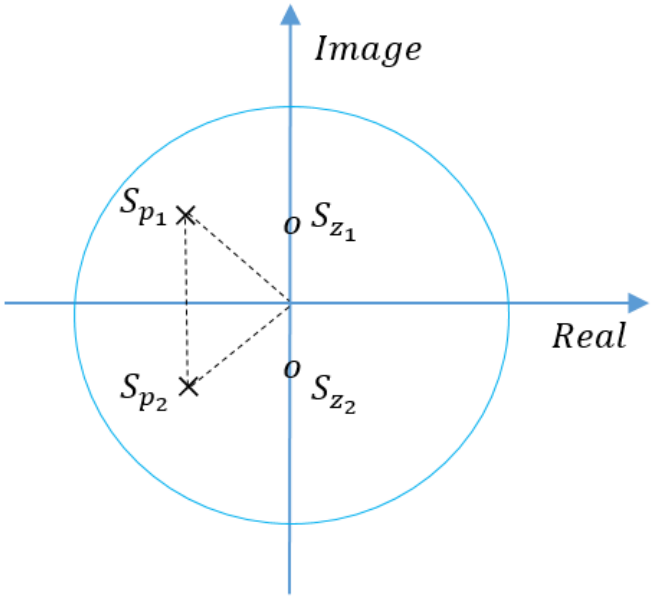


Figure 9. Poles and Zeros in the complex plane.

Zeros,

$$s^2 X_1 = 0 \quad (13)$$

$$\leftrightarrow s_{z_{1,2}} = \pm j\frac{2\pi k}{t_3 - t_2} \quad (14)$$

Poles,

$$s^2 + 2\varepsilon\omega_n s + \omega_n^2 = 0 \quad (15)$$

$$\leftrightarrow s_{p_{1,2}} = -\varepsilon\omega_n \pm j\omega_n\sqrt{1-\varepsilon^2} \quad (16)$$

The zero points together with pole points are illustrated in the complex plane as Fig. 9 below.

4. Planning the Proposed Scheme

In the present work, we discover the case of automated vehicle which lifting cargo to transport material in factory. The test scenario was set up as Fig. 10 which the proposed scheme is experimented on a model of vehicle. It is attached the flexible beam together with mass and an accelerometer above. In front of vehicle, one digital camera is hung on anterior cover. The distance to conflict with objects could be estimated since it returns all coordinates in workspace. The test surrounding is internal area of textile factory where there are a lot of workers. As a result, the potential accident between human and machine could happen at any time. The proposed scheme is able to overcome these troubles perfectly.

In detail, the automated guided vehicle is driven by using a TI's board to output PWM signal. A powerful DC driver that is capable to protect the over current or isolate the reverse current, is motivated to control DC motor. The accelerometer is a Sensors Booster-Pack, manufactured by Texas Instrument and include several sensors into one pack. We utilize this sensing module to recognize the magnitude of oscillation in respect to time. This sensor would direct data back to host server through UART protocol. Whenever the mobile vehicle moves, with mass, the oscillation occurs to act on aluminum beam. The swinging of beam is recorded by sensing module. Later, oscillating data is transmitted to host personal computer. Moreover, a Kinect camera locating ahead of robot, has fully functional image processing techniques. We program an interface software with camera in C/C++ language.

Theoretically speaking, it is desired that there exists an effective and feasible method to reduce the residual vibration in such situation. To verify our approach, the polynomial cubic spine, exponential spline of degree 3, exponential tension spline of degree 3 and exponential general spline scheme are recommended to achieve free-vibration motion. The geometric trajectory is collected via linear lines with circular lines in a sequence of equally segment due to control points p_i. Once, the flexible curves are generated to evaluate the exactly tracking ability of system and stable response. However, owing to residual vibration, the actual performance differs with command curvatures of the bends. For such fluctuation, it is traditionally considered that robot's direction movement along with y-axis would produce the residual vibration concerned on x-axis response. The higher velocity the autonomous robot reaches to, the larger amplitude of oscillation takes place.

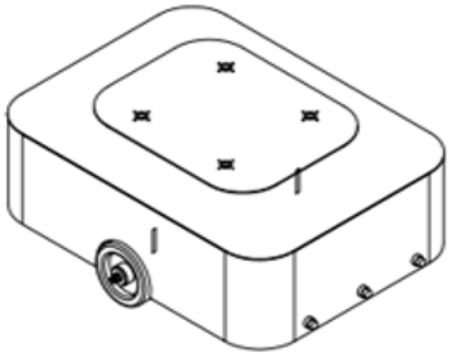


Figure 10. Simulation model of autonomous vehicle.

Based on proposed approach, an experimental prototype of potential collision has been built as Fig. 11. The differential drive mechanism depending on dissimilar speeds, orientates itself to turn left or right.

5. Discussion

The path planner based on cubic spline curve is denoted as Fig. 12. A series of B-spline basis functions are synthesized to perform the flexible and bended trajectory. Nevertheless, at each sampling time, there is only one basic function that is the most weighted scheme. Also, the determination of basic functions depends on control way-points. As we expect, the routing passes through all the given way-points.

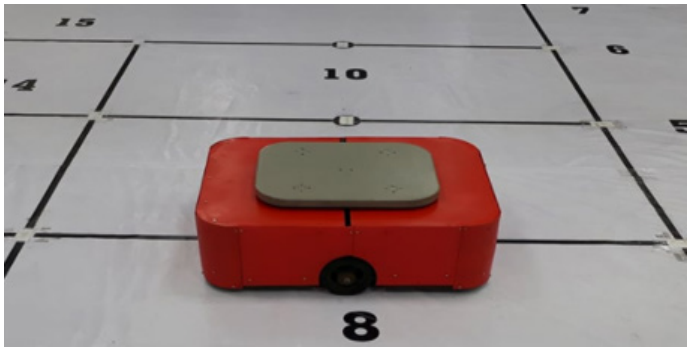


Figure 11. Practical model of autonomous vehicle.

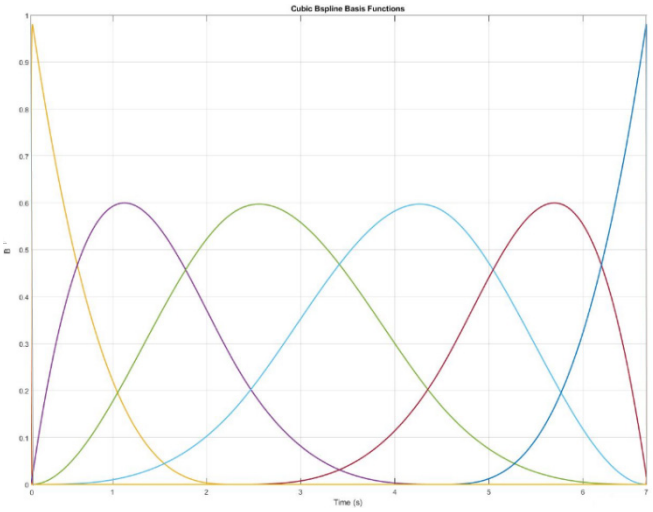


Figure 12. Cubic spline trajectory generation.

Note that there is a unique synthetization of control points which meets the way-point constraints. It could offer more degrees of freedom by swelling the number of control points. To utilize this generation more effective and real-time cost, a powerful CPU needs to be explored. The B-spline based motion profile of mobile vehicle is achieved as Fig. 13. The motion generator outputs driving command owing to interpolation functions. It provides an exact characterization of such category of routing line. While moving with feasible profile, the autonomous robot could ensure to reach desired parameters such position, speed, acceleration and jerk. To measure the impact of proposed approach, the reaction of mass m_2 is recorded in Fig. 14. The variation of m_2 tends to zero when time is infinite. From these results, it can be considerably seen that the path generation using B-spline strategy in respect to analysis of dynamic characteristics has been developed successfully. The wide range of industrial applications are recommended this study in future.

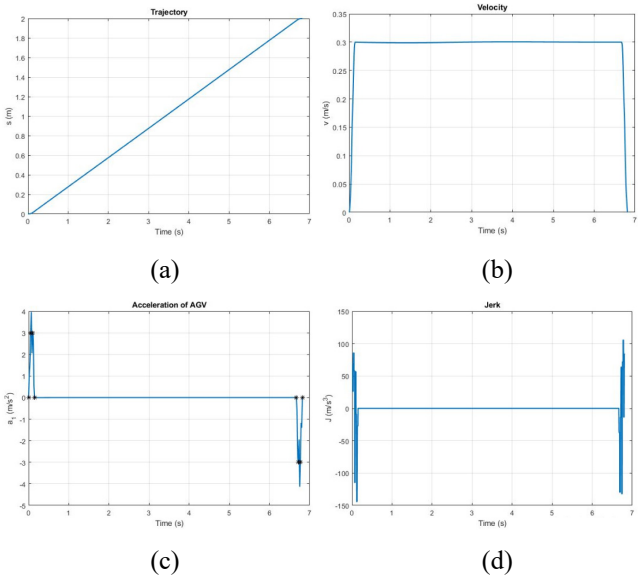


Figure 13. Motion parameters of autonomous vehicle, (a) traveling distance, (b) velocity, (c) acceleration and (d) jerk.

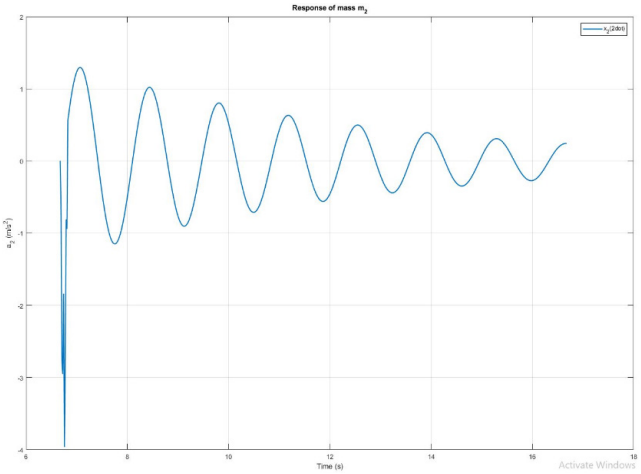


Figure 14. Actual response of carrier m_2.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] X. T. Truong and T. D. Ngo, "To Approach Human?: a unified framework for approaching pose prediction and socially aware robot navigation," IEEE Transactions on Cognitive and Developmental Systems, vol. 10, no. 3, pp. 1-17, 2018, https://doi.org/10.1109/TCDS.2017.2751963.

[2] F. Yang and C. Peters, "Social-aware navigation in crowds with static and dynamic groups," 11th International Conference on Virtual Worlds and Games for Serious Applications, pp. 1-4, 2019, https://doi.org/10.1109/VS-Games.2019.8864512.

[3] W.E. Singhose and N.C. Singer (1996), "Effects of input shaping on two-dimensional trajectory following," IEEE Trans. on Robotics and Automation, 12(6), https://doi.org/10.1109/70.544771.

[4] Q. T. Truong, H. Q. T. Ngo, T. P. Nguyen, H. Nguyen and W. H. Kim," A Novel Infrastructure Design of Industrial Autonomous System," International Journal of Fuzzy Logic and Intelligent Systems, 19(2), pp. 103-111, 2019, http://doi.org/10.5391/IJFIS.2019.19.2.103.

[5] H. Q. T. Ngo, Q. C. Nguyen, W. H. Kim, " Implementation of Input Shaping Control to Reduce Residual Vibration in Industrial Network Motion System," 2015 15th International Conference on Control, Automation and Systems, pp. 1693-1698, 2015, http://doi.org/ 10.1109/ICCAS.2015.7364629

[6] L. Biagiotti and C. Melchiorri (2012), "FIR filters for online trajectory planning with time- and frequency domain specifications," Control Engineering Practice, vol. 20, no. 22, pp. 1385-1399, 2012, https://doi.org/10.1016/j.conengprac.2012.08.005.

[7] H. Q. T. Ngo, T. P. Nguyen, T. S. Le, V. N. S. Huynh and H. A. M. Tran, "Experimental Design of PC-based Servo System," International Conference on System Science and Engineering, pp. 733-738, 2017, https://doi.org/10.1109/ICSSE.2017.8030973 .

[8] L. Biagiotti, C. Melchiorri and L. Moriello (2015), "Optimal trajectories for vibration reduction based on exponential filters," IEEE Trans. on Control Systems Technology, vol. 24, no. 2, pp. 609-622, 2016, https://doi.org/10.1109/TCST.2015.2460693 .

[9] H. Q. T. Ngo and M. H. Phan, "Design of an Open Platform for Multi-Disciplinary Approach in Project-based Learning of an EPICS Class," Electronics, 8(2), 200, https://doi.org/10.3390/electronics8020200.

[10] L. Biagiotti and C. Melchiorri, Trajectory Planning for Automatic Machines and Robots, Springer-Verlag Berlin Heidelberg, 2008.

[11] H. A. M. Tran, H. Q. T. Ngo, T. P. Nguyen and H. Nguyen, "Implementation of vision-based autonomous mobile platform to control by A* algorithm," International Conference on Recent Advances in Signal Processing, Telecommunications and Computing, pp. 39-44, 2018, http://doi.org/10.1109/SIGTELCOM.2018.8325802.

[12] N. M. Phu, V. Tuyen, T. T. Ngo, "Augmented heat transfer and friction investigations in solar air heater artificially roughened with metal shavings," Journal of Mechanical Science and Technology 33, 3521-3529 (2019), http://doi.org/10.1007/s12206-019-0646-x

[13] L. Biagiotti and C. Melchiorri (2011), "Input shaping via b-spline filters for 3-d trajectory planning," In IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, http://doi.org/10.1109/IROS.2011.6094836.

[14] N. M. Phu, N. T. M. Trinh, "Modeling and experimental validation for off-design performance of the helical heat exchanger with LMTD correction taken into account," Journal of Mechanical Science and Technology 30, 3357-3364 (2016), http://doi.org/ 10.1007/s12206-016-0645-0

[15] H. Q. T. Ngo, Q. C. Nguyen and W. H. Kim, "Implementation of input shaping control to reduce residual vibration in industrial network motion system," International Conference on Control, Automation and Systems, pp. 1693-1698, 2015, http://doi.org/10.1109/ICCAS.2015.7364629.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 100-105 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Implement Wireless and Distributed Vibrator for Enhancing Physical Activity of Visually Impaired Children

Hung-Chi Chu[1], Fang-Lin Chao[*,2], Liza Lee[2], Pei-Yun Kao[2]

[1] *Department of Information and Communication Engineering, Chaoyang University of Technology, 436, Taiwan R.O.C.*

[2] *Department of Industrial Design,* [3]*Department of Early Childhood Development & Education, Chaoyang University of Technology, 436, Taiwan R.O.C.*

ARTICLE INFO

Article history:
Received: 27 October, 2019
Accepted: 08 January, 2020
Online: 15 January, 2020

Keywords:
Tactile sensation
Bodily activities
Visually impaired
Music
Wireless network

ABSTRACT

Learning and mental and physical counselling for visually impaired children are important. However, due to the lack of visual sense assistance, Physical Activity will gradually degenerate without proper enhancement, which may require the care of others for life. Inner fear and sensitive psychological factors hinder the development of living ability. Strengthening physical activity of visually impaired children improves their ability and reduces psychological obstacles, which is conducive to getting along well with others. The distributed vibration developed in this study can enhance children's physical exercise. Through music rhythm, children's psychological resistance is reduced, and their body movements are triggered by distributed vibration, which has an impact on the physical activity of children with visual impairment.

Seven kindergarten-age children were selected to participate in a 16-week tutoring. A score was recorded based on the assessment form, teaching record, and interviews with parents or students. The results showed that vibration and music are helpful in developing a young child's motor skills. After two sessions of tutoring, player became more interested in participating along with the music and tactile signals. Players were able to correctly identify the limbs that corresponded to the vibration and gesture with their hands and feet. Assistive technology had a positive effect on the physical activity of the visually impaired children.

1. Introduction

The daily life of the blind is not as active as ordinary people. The lack of visual sensory assistance results in problems such as slow movements and inability to respond immediately to the surrounding environment. Failure to integrate into normal life causes personality shrinkage, lack of contact with people, and decreased perception. With the advancement of technology, caring assistive technology has been applied to assist disadvantaged groups, such as: smart cane, route guidance, and voice/ image recognition. There were some facilities regarding to motor learning in visually impaired children [1,2,3]. Anthierens utilized a feel space navigation belt composed of 16 buzzers to distribute sensory cues for visually impaired kayakers [1]. Per Ellis presented sound therapy is effective for educating children who have multiple learning difficulties and may provide additional ways of communicating [4].

1.1. Haptic Devices & Teach

Haptic devices which interact with the user without disturbing them are most commonly used for drawing the attention of cell phone owners. Small direct current (DC) motors offer high performance and a quick response time for a low price. Piezoceramics convert electrical energy into mechanical vibrations. Anvik [5] had investigated the potential of rich tactile notifications with six vibrators.

Previous studies have also shown positive effects through airflow and tactile sensations [6] in promoting children's physical activity. Baker et al. developed a 'ring' to be worn by an orchestral conductor and a 'vibration matrix' through a two-dimension haptic signal through which the conductor could convey their intent to a visually impaired musician [7]. Low-cost vibrators had also been constructed to promote general distribution of the application to disadvantaged group [8].

* Fang-Lin Chao, Email: flin@cyut.edu.tw

1.2. Purpose

The aim of this study was to examine the effect of distributed vibration in enhancing the motor control of visually impaired children with disabilities. Both quantitative and qualitative methods were utilized to measure the validity of the intervention. Next, six second-grade elementary school students participated in an activity which paired vibration events with song segments. During the session, players were able to correctly identify the limbs that corresponded to the vibration and gesture with their hands and feet.

Design actuators and tactile devices supported the effectiveness of distributed vibration in enhancing the movement of children. The specific purposes of the study were:

(1) To design affordable equipment for enhancing exercise behaviors.

(2) To examine the effectiveness of the equipment of distributed vibrator and music activities in enhancing physical activity in children.

2. Assistive Device Design

2.1. Concepts and circuit implementation

The initial concept art of the vibrator application is shown in Figure 1. This concept art proposed tiny distributed vibrators to prompt movement at specific positions on the user. Tactile signals cue the user to respond with an action at an individual location.

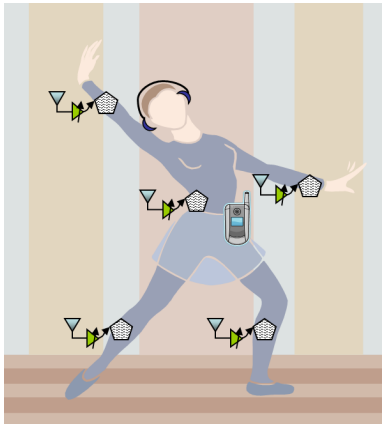


Figure 1: Concept art of distributed vibrators

For proper tactile cueing, the vibrators must contact the skin. In the early stage of the field test, we used elastic straps to attach the vibrator to the corresponding limb. Vibrators of the appropriate size received a signal through a wire; a computer controlled the on/off timing of each channel.

The prototype's circuit consisted of solid-state relay (SSR) components through a USB line. The control channel employed a Phidget Interface Kit 8/8/8 I/O board, which controlled the switching by preset sequence. An external DC power supply with solid-state relay was constructed (Figure 2) with a 360 ohm resistor in series with a 3.6V Zener diode to ensure the current amount.

Figure 3 shows the graphical user interface (GUI) of the control program. The interface was developed using Microsoft Visual Basic, which includes five functions. The first part is the tool bar to provide the connection, COM part setup, Record, Stop, Execute, Camera, and Exit functions. The second part is the melody loading, and the third part is the load/modify function of body movements, which can load preconfigured body movements. Users can choose specific vibrator through push button: LH (left hand), RH (right hand), LF (left foot), and RF (right foot) buttons to active movements. Finally, parts four and five show the body movement results.

After selecting the music, users pressed the top button to begin. The time and order of the button pressing were recorded in a text file for playback during the next test.

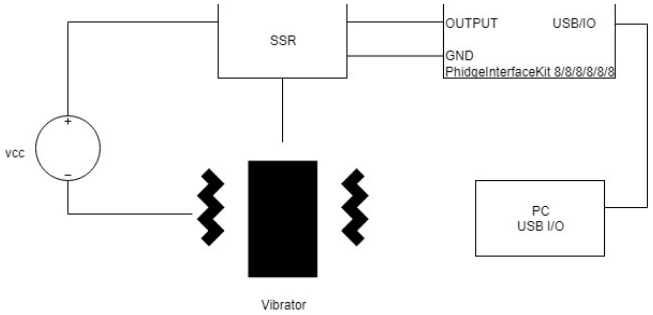


Figure 2: Hardware block diagram of the solid-state relay and vibrator

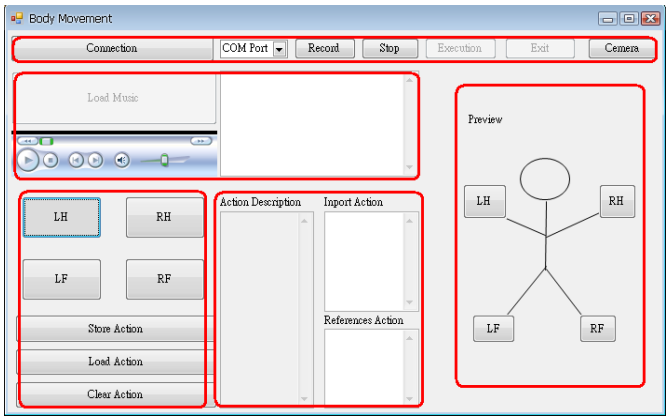


Figure 3: Editor screen layout of vibrators and control.

2.2. Wireless WSN implementation

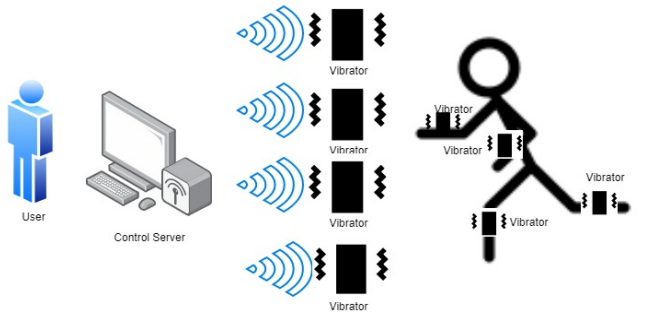
In addition, with the rapid evolution of telecommunications technology, assistive technology combined with microelectronics and information technology has benefits. Wireless sensor network is an infrastructure less, low cost, dynamic topology, application oriented design with wireless distributed nods. WSN designing become more complex for further processing [9]. Deployed sensors provided the required connectivity [10]. Sensor and the wireless implementation of the distributed vibrator are reviewed in this section.

A WSN emphasizes low-cost, low-speed, and ubiquitous communication between devices. The basic framework involves a 10-m communication range with a transfer rate of 250 kbps. The method significantly reduces line interference during movement. The system (Figure 4a) shows a central server using a ZigBee network interface, which is responsible for transmitting control signals and receiving messages from sensor nodes with the vibrator. The control command could turn on/off the vibrator with a specific time (Figure 4b). Each sensor module uses a 4.2 V lithium battery power. The vibrator is connected to a low-speed I/O pin on the sensor node to reduce the module's energy

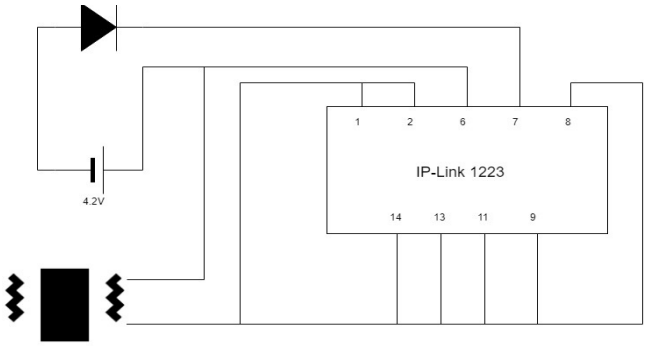
consumption — the hardware circuit prototype placed within a chassis made by 3D printing (Figure 5). The ZigBee interface is IP-Link5501, and the sensor node is IP-Link1223 [11].

Zigbee's Network Layer Protocol Data Unit (NPDU) shows in the Figure 6. An NPDU consists of a network header and a network payload. The Network header contains frame control of 2 octets, routing fields of 6 octets, and data payload NSDU (network service data unit) of variable length. Besides, the routing fields of 6 octets are composed of a destination address of 2 octets, a source address of 2 octets, a radius (broadcast radius) of 1 octet, and a sequence number of 1 octet. Among them: the destination address of 2 octets is set to an address of 0xFFFF to indicate broadcast information, thereby broadcasting the control signals of the four vibrators at the same time.

The sequence number of 1 octet is the broadcast sequence number (BCSN). The control field of the vibrator used in this research is to use the 1 byte in the original data payload to mark it as ID1, ID2, ID3, and ID4. Each ID represents a controlled vibrator, and its field length is 2 bits, in which the bit value code: 11 means intensity vibration, 10 means moderate vibration, 01 means low vibration, and 00 means stop. ID1 and ID2 represent the left, right-handed vibrator, ID3 and ID4 represent the left, right-foot vibrator.

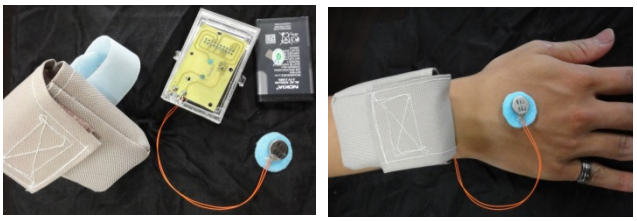


(a)



(b)

Figure 4: (a) Architecture of the assisted body movement system, (b) the circuit of vibration node.



(a) (b)

Figure 5: (a) Vibration node prototype with ZigBee Interface (IP-link 5501) and Sensor node: (IP-link 1223) (b) hand wearing.

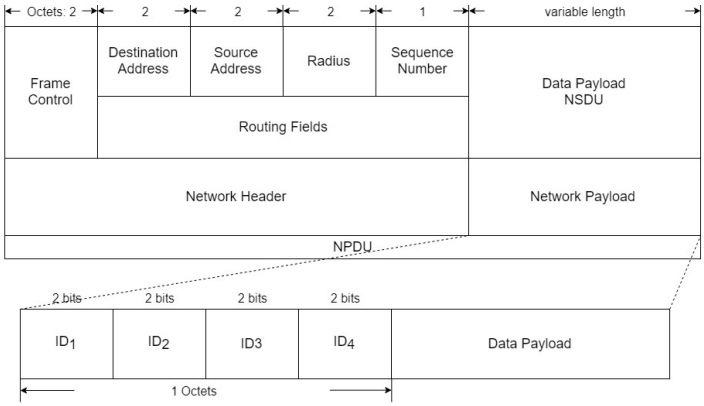


Figure 6: Network Layer Protocol data unit of Zigbee.

The increase of the volume of the module weakens the vibration amplitude, which affects the perception and response of the subject. The designer tried EVA and placed vibrator under the main circuit with an eccentric shaft. The main circuit is separated from the round flat vibration motor. The arc shape conforms to the curvature of the wrist, and the fabric and the devil felt are connected for convenience (Figure 7).

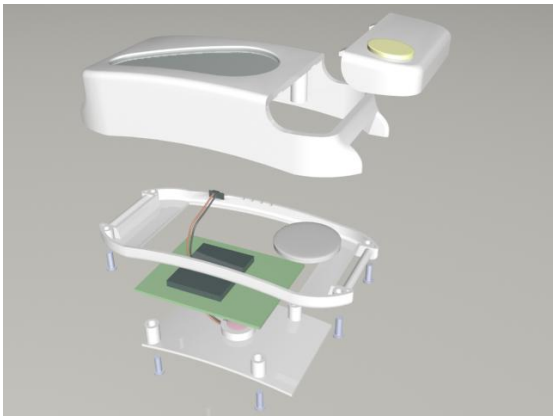


Figure 7: Module consists of an upper cover, chassis, and a battery cover. The chip is located on the lower chassis. There are power switch, mode setting button, and LED display switch.

2.3. Human Factor issues

Human factors design prompted ergonomic considerations for specific users. An interdisciplinary approach consist human factors, ergonomics, and work psychology in connection with design goals for cyber-physical product [12]. We kept the total weight of each vibrator under 130 grams through circuit miniaturization. The elastic straps followed the contours of the arm, wrist, or leg. Human factors were measured data. The maximum wrist circumference of the male wrist is about 19.8 cm and the average circumference of the adult ankle is about 25 centimeters. The initial design, featuring silicone material to help maintain skin contact with the backside-mounted vibrator, is shown in Figure 5.

The vibration amplitude/frequency need adjust for the specific user. Primary parameters of perception are: amplitude, frequency, timing, and location [13]. The skin is roughly sensitive to vibrations between 20 and 500 Hz. Vibration stimuli exceeds threshold will be detected. The threshold for the trunk is 4 microns at 200 Hz [14, 15]. High spatial resolution located on the hands, and face. Tactile displays must be setting between threshold and the maximum comfort level. Comfortable stimuli range 15-20 dB

above the absolute threshold. Amplitudes above 0.6-0.8 mm will result in a pain sensation. Design considerations reviewed considering: frequency, intensity, area of stimulation [16].

The stimulation responses of retarded children were analyzed in [17].Vibration sensation at low and high frequencies differed significantly. Sense thresholds lower frequencies (16 Hz and 31.5 Hz) were low. The relatively unchanged tactile frequency discrimination between vibration frequencies of 180 and 250 Hz were measured in [18].The movement of the device limited to a narrow bandwidth of frequencies below 30 Hz [19].

Devices for locating or stimulating reflex points in the body with alternating magnetic fields producing oscillating movement which preferably exerting perpendicular forces on the skin [20]. Children's availability and movement convenience were prioritized for location choosing. When setting the vibrator, avoid high frequency or intensity to prevent tactile shock (Figure 8).

Figure 8: Wired module and external power supply.

3. Methodology

Visually impaired children have different athletic abilities with age and innate physical conditions. We have arranged tests for two ethnic groups:

(1) Children with multiple impairments of varying degrees of severity, aged between 3-6 years old, have limited ability to autonomously move. Due to the combined effects of other obstacles, the ability to move is limited, and the personality is closed and passive. Physical exercise encourages them to try to show more variety or amplitude.

(2) Visually impaired children have normal ability - aged 9-12. They are amblyopic or totally blind, but their physical capabilities are close to the average person. Close observation reveals that students have limited ability to balance, with small and restrained movements. Usually, they are accompanied by each other and passively do the activities within the allowable range under the teacher's request. Students may have obesity due to less activity. Physical exercise encourages them to demonstrate amplitude rhythm and interaction.

Based on the purpose of the study, the specific research questions asked were:

(1) Could multiple vibrators and music activities assist a young child with visually impaired development of motor skills?

(2) Could multiple vibrators enhance a visually impaired child's range of movement?

3.1. Wired test

The wired module is first introduced, it had limitation of connected wire but the facility cost is low and easy to implement. Teachers can give exercise tips in a small area. Two phases of testing were arranged to understand the adaptation to the vibrator for different ability: Phase- I involved group (1) 3-6 years-old with multiple disable, and Phase- II anticipate group (2) 9-12 years-old with normal ability.

Quantitative analysis utilized measuring validity based on assessment scales, semi-structured reports, and video recording. The therapeutic procedures included the establishment of goals, assessment, and activity evaluation. Some teaching styles were chosen based on participants' learning responses. The study took place at a school in central Taichung, Taiwan. Seven visually impaired participants were selected. Three participants exhibited severe physical disability, and the other four children exhibited moderate physical disability. The Phase-I participants were given 40-minute tutoring session once per week for 16 weeks (Figure 9).

Figure 9: Field practice during Phase-I tutoring

The quantified results of a semi-structured musical activities observation form were compared with related background documents (reports from the parents and teaching logs). The forms were completed by a physical therapist in order to gather physical movement data which was scored by three observers.

3.2. Wireless test

The wireless module lifts the distance limitation of the vibration connection. In addition to prompting in a small range, you can also provide motion prompts in a large area covered by radio waves. The wireless module test also carried out to understand the adaptation to the vibrator for different ability group. The approaches were similar to wired test except utilized the wireless modules.

4. Results and Discussion

4.1. Results of wired test

Phase- I Observation

Figure 10 shows the participants' physical movements during the test. After attending the program, all participants exhibited significantly improved physical mobility. The three participants

with severe symptoms were able to move their heads and hands slightly. The other four participants with moderate symptoms were able to walk and run stably and change direction. The tutoring sessions had a positive result (Figure 11).

Week 1 to Week 4: At the beginning of the research, none of the participants had the motivation to perform the movements. When the instructor put the distributed vibration equipment on the participants' limbs, and did not know what to do. The four participants with mild to moderate symptoms tried to pull the attached accessory. The frequency of the vibration was high and fast, consequently the participator could not follow it accordingly. A preset guidance protocol was required

Week 5 to Week 11 (same experiment setup): After the team adjusted the frequency of the vibration signal, most of the children did not reject the vibration and began to move to the music. Those with mild and moderate symptoms were able to move when participator felt the vibration and heard music. Sometimes participator would tell the instructor "felt the vibration on skin". The researchers found that the fuse of the attached accessory and cables was too long; as a result, the vibration signal did not transmit properly sometimes. Therefore, the team improved the cable routing and electrical contact.

Week 12 to Week 16: All the participants made progress at different levels (Figure 11). The participants with severe symptoms showed facial expressions of happiness. The participants with less severe symptoms were able to respond to the vibration and music correctly. The performance is slowly increasing by the time not right after using this equipment. For the mental state of the visually impaired with multiple obstacles, they couldn't accept changes from outside quickly. People with multiple disabilities have weak movements and expressions. After getting used to tactile sensations and being able to respond appropriately, additional channel of communication prompts them to participate more.

Figure 10: Participants' physical movement with (a) severe and (b) moderate symptoms

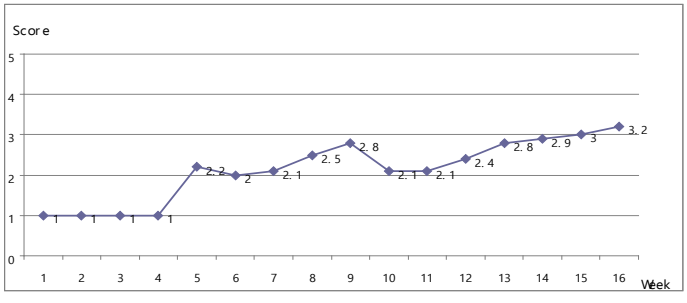


Figure 11: Observers' assessment of participants' physical movement

Phase- II Observation

Vibrators were placed on the upper and lower body (subjects' hands, wrists, and ankles). Six elementary school students participated in a five-minute session where vibration events were paired with segments of a song. The vibrators were turned on, along with the popular song, "Sorry," by Super Junier. Subject C01 (Figure 12a) appeared initially frightened and covered her ears. After the teacher's guidance, C01 was able to complete the directed actions.

(a) (b) (c)

Figure 12: (a) C01's reaction, (b) C02, (c) C04/05situation

At first, Subject C02 (Figure 12b) found the experience strange and a little frightening. However, after guidance, C02 became more relaxed and willing to jump in place. Subject C03 swung her body around and occasionally moved her wrist in small motions. After some instruction, Subject C03 was able to wave her left arm. Subject C04 was more lively, swinging his body rhythmically and creating his own dance moves. He shook his body during the interlude and appeared to be having fun. Subject C05 (Figure 12c) could correctly identify which vibrator was vibrating and raise and lower his foot, but his rhythm was not as good as C04's. Both participants corresponded to the vibrations and made appropriate gestures with their hands and feet. Exercises make them more aware of their relative positions.

4.2. Results of Wireless usage

Ten special children are invited to perform small-scale sports, including children with low vision, blind children, and slow intellectual development. After instructing the child to equip the child with a vibrator, observe the response with the music and understand the differences in the way of using it.

Figure 13: Children spontaneously tap the instrument

Children spontaneously tap the instrument according to the vibration instructions. Special children feel keen, and some children want to take off the vibrator because their hands are not comfortable. A child with a slow intellectual development has an absolute sensation of music show outstanding performance. Two

student laugh and very focus on the indicator and tap the instrument. Some children complete the test through the guide of the accompanier. (Figure 13)

5. Conclusions

Through music activities with children of varying ages, the findings of the study were as follows: The participants' capacity for physical movement improved significantly. We also observed an improvement in social validity from the teachers and participants. The tactile design was able to assist the development of the participants' physical skill. The wireless module lifts the distance limitation of the vibration connection. In addition to prompting in a small range, you can also provide motion prompts in a large area covered by the radio waves.

Vibrator devices reinforce active participation for children with severe disabilities. Students exhibited high willingness to withstand new stimuli. Therefore, it is necessary to adapt a teaching strategy to the characteristics of the students. Wireless technology enables remote instructions to be given to assist users in interacting with the people and environment. This is not something that can be achieved by purely hardware design. The effectiveness of the equipment design also relies on developing the teaching strategy and events.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported in part by the National Science Council, Taiwan, ROC, under grant NSC 101-2221-E-324-028, and 99-2221-E-324-026-MY2.

References

[1] Anthierens, Cédric, Didier Groux, and Vincent Hugel, "Sensory navigation guide for visually impaired sea kayakers" Journal of Field Robotics, **35**(5), 732-747, 2018.
[2] Zuzanna Maćkowiak, Wieslaw Osiński,"The effect of sensorimotor training on the postural stability of visually impaired women over 50 years of age" Journal of Women & Aging, 27(1), 68-80, 2015. DOI: 10.1080/08952841.2014.928140
[3] Malik, S. et al., "Orientation and mobility training in special education curriculum for social adjustment problems of visually impaired children in Pakistan" International Journal of Instruction, 11(2), 185-202, 2018.
[4] Phil Ellis, "Incidental music: A case study in the development of sound therapy," British Journal of Music Education, 12(1), 59-70, 1995.
[5] Cecilie H. Anvik, "Embodied spaces in the making: visually impaired people, bodies and surroundings" Scandinavian Journal of Disability Research, 11(2), 145-157, 2009. DOI: 10.1080/15017410902830710
[6] Schwellnus, H., et al. "Using movement-to-music technology for play with children with special needs" OT Now, July (2002).
[7] Baker, David, Ann Fomukong-Boden, and Sian Edwards, "'Don't follow them, look at me!': Contemplating a haptic digital prototype to bridge the conductor and visually impaired performer" Music Education Research 21(3), 295-314, 2019.
[8] Bischof, A. et al. "Exploring the playfulness of tools for co-designing smart connected devices: a case study with blind and visually impaired students" Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion,. ACM, 2016.
[9] Jangra, Ajay. "Wireless sensor network (WSN): Architectural design issues and challenges" 2010.
[10] Mohamed, Shaimaa M., Haitham S. Hamza, and Iman Aly Saroit, "Coverage in mobile wireless sensor networks (M-WSN): A survey" Computer Communications, 110, 133-150, 2017.
[11] IP-Link 1223, Helicomm Inc., http://www.helicomm.com/showproducts.asp?ID=144
[12] Stern, Hendrik, and Till Becker, "Development of a model for the integration of human factors in cyber-physical production systems" Procedia Manufacturing, 9, 151-158, 2017.
[13] Van Erp, J. B., "Guidelines for the use of vibro-tactile displays in human computer interaction" In Proceedings of eurohaptics, Vol. 2002, 18-22, 2002.
[14] Verrillo, R.T., "Effect of contactor area on the vibrotactile threshold" Journal of Acoustical Society of America, 35 (12), 1962-1966, 1963.
[15] Sherrick, C.A. & Cholewiak, R.W., Cutaneous sensitivity. In Boff, K.R., Kaufman, L. & Thomas, J.P.:Handbook of perception and human performance (pp. 12-1 –12-57), New York: John Wiley and sons, 1986.
[16] Verrillo, Ronald T. "Psychophysics of vibrotactile stimulation." The Journal of the Acoustical Society of America 77.1 (1985): 225-232.
[17] Pujol, Kelley Kruse. "The effect of vibrotactile stimulation, instrumentation, and precomposed melodies on physiological and behavioral responses of profoundly retarded children and adults" Journal of Music Therapy, 31(3), 186-205,1994.
[18] Harada, Noriaki, and Michael J. Griffin. "Factors influencing vibration sense thresholds used to assess occupational exposures to hand transmitted vibration" Occupational and Environmental Medicine, 48(3), 185-192, 1991.
[19] Wall, Steven A., and William Harwin. "A high bandwidth interface for haptic human computer interaction." Mechatronics 11.4 (2001): 371-387.
[20] Herman, Goldfarb. "Massaging garment with vibrators located in back and chest sections" U.S. Patent No. 3,310,050. 21 Mar. 1967.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 106-111 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Comparative Study of Semantic and Keyword Based Search Engines

Bzar Khidir Hussan *

Department of Information System Engineering, Erbil Polytechnic University, Erbil- Kirkuk Road, Hadi Chawshli Street- Kurdistan Region- Iraq

ARTICLE INFO

Article history:
Received: 30 September, 2019
Accepted: 28 December, 2019
Online: 15 January, 2020

Keywords:
Conventional Search Engines
Syntactic Search Engines
Semantic Web
Intelligent Engines
Semantic Search Engines

ABSTRACT

Day by day, the data on the web becomes very huge which makes it difficult to find relevant information. Search engines are one of the successful factors that can retrieve information from the Web. The process of seeking information by search engines helps users find information on the internet, however it is not an easy task to find the exact information from this massive data available on the Web. Semantic Web technology has an ability to focus on metadata rather than syntax, which made the semantic search engines to search for the meaning of keywords instead of the keyword syntax. Consequently, an effective role of performance in conventional search engines can be achieved by rising the accuracy of information returned by a search query. In this paper, a survey for syntactic-based search engines and semantic-based search engines are studied, a comprehensive comparison between the two is presented, finally, their technologies are compared and discussed.

1. Introduction

The term Web is one of the most important technologies that allow users to access huge and different information through different locations in the world [1]. This information which is stored on servers is typically unstructured or - semi structured data [2,3]. The data on the Web is tremendously increasing which leads to many obstacles such as difficulty of finding relevant information or discovering exact knowledge on the Web [4].

The first generations of Web search engines, such as AltaVista, were indexing the contents of Web pages. The second generation of search engines, for instance Google, were considering the links to/from a Web page as a method of determining relevance. Both generations were mostly syntactic, which means they were depending on the keywords as text in their queries. Searching for interested information in most of the current search engines will result of retrieving hundreds of thousands retrievals while most of them are not relevant to what is meant to be found [5,6]. Hence, to solve such difficulties, Semantic search is intended to be updated so that it depends on the meaning of the keywords instead of the context [7].

Semantic Search Engines (SSE) use Semantic Web (SW) technology in their systems which make them intelligent to retrieve related information on the Web [8]. The SW aims at providing information as formal, well defined meanings, compatible, sharable knowledge base, and can be processed by machines [3]. Ontology acts an important role in the SW technology as it's famous as of the backbone of the SW structure, and is the vital element of SW infrastructure [7]. Web Ontology Language (OWL) and Resource Description Framework-Schema (RDFS) are the recommendations of the World Wide Web Consortium (W3C) for data representation models so as to deliver foundations for the ontology descriptions [4]. Ontologies provide distinct descriptions in their information, as a result, they are used in numerous fields and applications since its knowledge representation is understandable and processable by software agents and systems [8]. Ontology is a collection of semantically related concepts built on a limited number of predefined relations and terms of a domain. These terms and concepts can be represented visually so as to ease the representation for both syntactic and semantic data [9]. In Web, once abstract data is distributed across several knowledge bases, ontologies are the solely resolution as commonplaces to interpret the mutual senses of the domain key terms. Hence, significant concerns seek the development of ontologies [10].

Transforming syntactic search engines to semantic ones is not an easy task, since in the later one search results rely on the meaning of the query keywords, henceforth the search engines have to understand the keywords semantically in order to

* Bzar Khidir Hussan, Bzar.hussan@epu.edu.iq

retrieve relevant information. To do so, a new layer could be added to the so called syntactic search engines [11]

This paper presents the Web and SW technologies along with the Keyword and Semantic search engines. Keyword Search Engines (KSE)s are presented in section 2, section 3 reveals SSEs. In section 4, technologies of SSEs are presented, and section 5 exposes the most common SSEs. In section 6, a comprehensive study to the literature is given and a comparison of the works is presented. Section 7, gives a discussion panel for the studied systems. Finally, the suggestions and conclusions finalizing the research.

2. Keyword Search Engines

Conventional Search Engines are very helpful in finding information on the internet and getting results within some time, but they suffer from the fact that they do not know the meaning of the terms and expression used in the web pages and the relationship between them [12,13]. Surveys show that users who seek for information on the web do not find accurate results in the first set of URLs returned, because of increasing size of links on the web pages. Sometimes one word has several meanings and several words have the same meaning, in that case if a user wants to search for a particular word then it may produce confusion and user will not get what he wanted to search [2,14].

Search Engine Optimization (SEO) are used to find and search alternative search terms that people use with search engines while looking for similar subjects [15]. SEOs search for more keywords, which are used to achieve better rankings in search engines. Once a different keyword is found, they expand on it to find similar keywords for that keyword [16]. Keyword suggestion tools usually aid the process of finding similar keywords such as in [17] where substitutional keywords are the suggestions for the query. There are many techniques used in keyword search engines such as identify the core of the keyword, research related search terms, create a list of main terms and long-tail keywords, use the Google AdWords keyword planner…etc. [18].

3. Semantic Search Engines

The Semantic Search Engines (SSE)s are the intelligent engines that search for keywords depending on their meaning [19]. In addition, they guarantee the results those are related to the meaning of searched keywords. SSE use ontologies so that they achieve the meaningful retrievals and get a high accuracy result. [12]. SW considers as a Web 3.0, or an extension to the current web which represent information in order to link information in the web as a form of HTML, OWL and RDF files [5]. The SSEs are distinguished to have several types of relational links among verity types of resources instead of the single relation of resources. There are many examples of SSEs such as Hakia, DuckDuckGo, Swoogle… etc. The methods to store information within the SW technology are able to answer complex queries given to a search engine [17][12].

3.1. Technologies of Semantic Search Engines

The SSEs are essentially based on some technologies or methods which effectively achieve the SSEs [20]. These technologies are sometimes called the SW layers which include applying some tools such as inference engines, rule languages, annotation tools… etc. One of the main technologies used in the SSEs are the ontologies which can be offered within the form of RDF, RDFs, and OWL [21]. These technologies are used in the structure of SSE which briefly can be described below:

3.2. Unicode and URI

Defines as the base level of the SW technology which is used for identification and determination of the resource location. The Unicode is used to standardize the letters of machines, while the URI is used for identifying each resource by unique name [22].

3.3. Extensible Markup Language (XML)

A subset of Standard Generalized Markup Language (SGML) and machine readable represented by markup language. This language is widely used in the web for some reasons such as it is simple, flexible text format and its structure used to describe data. XML meets the challenges of E-business and electronic publishing as well. In addition, it has a very important role to exchange between different kind of data on the web [23].

3.4. Resource Description Framework (RDF)

It is representing the primary layer used in SW. RDF is very important to represent data which can be processed by machines [17]. The method that is used to identify and provide the relationship among the resources called graph model. The best simple model language which is used in this layer is RDF Schema which is used to create relations and descriptions of resources [24].

3.5. Ontology Vocabulary

The ontologies are used to describe data of the SW and improve methods to give uniform way to make easy communications among different parts of resources and be understandable by each other. Ontology is the method that can provide common grammar and vocabulary of data that are published, specifically the description of semantic data which represent by ontology [18].

3.6. Logic, Proof and Trust

Are the last layers of the SW cake which follow the ontology layer. These layers are used to check and solve consistency problems and the trustworthiness of SW. In addition, redundancy of concept duplicate data [25].

3.7. Common Semantic Search Engines

The SW technology has grown a new generation of the web by using some new methods to search about the best results related to searcher intent. There are many engines which depend on the semantic approach [17]. In the following sections, the most common SSEs are presented.

3.8. Hakia

It considered one of the common SSEs which provide results that are relevant to the concept of words instead of main

keywords. In other side, it's not just depending on the keywords but use the concepts of entire phrases such as questions or sentences. The one of the most important characteristics of Hakia is its capability to provide the results depending on equivalent concept such as "cure=treat" or "cons=disadvantages" [12]. The results of the search are divided into classes such as Web, News, and Video etc.; also, they can be divided according to the date or relevance. The technology which is used in Hakia is the OntoSem technology, which is a linguistic database [19]. The words here are categorized into different senses by depending on the QDEX (Query indexing technique), which considered as an infrastructure to index the data by Semantic Rank algorithm. This algorithm use the ontology and fuzzy logic to search about all possible requests [18].

3.9. Kngine

The Kngine divides the search results into two types: documents or images. This engine searches for the information related to the search term which means search about the concepts of the words [12]. It is very intelligent engine because its retrievals related to natural of question. For example, search about the city, the expected results will be related to the city lactation, events, weather and history [20].

3.10. DuckDuckGo

It has many of features that distinguish of other SSEs. When a keyword is being searched the results will have many related retrievals, meaning that, it provides different answers for the searcher, the searcher can choose the answer that is related to his intent [17]. For example, when we search about the term Apple, the engine will provide many of answer such as fruit, computer, bank, etc. [26]. The DuckDuckGo is distinguished among other types of semantic search engines by dealing with all users the same results when search the same term. Also, it deals with many other websites including Wikipedia, Bing and yahoo [27].

Table 1: Main Differences Between Keyword and Semantic Search Engines

No	Semantic search Engines	Keyword Search Engines
1	An approach to achieve a related and accurate information to user queries [12].	Considered as traditional engines which return the results depending on context of queries [18].
2	It depends on the stop words and punctuation marks. These marks effects on the search results [16].	It does not depend on the stop words and punctuation marks and the results are not accurate [21].
3	The OWL and FDF languages are the base languages used to creating web pages [19].	The HTML, XML and CSS languages are the base languages used for creating the web pages [12].
4	Seeks to provide the accuracy in returned information by understanding the meaning of keywords related to what the seeker desires [20].	It is searching exactly depending on words in the website which determined by searcher [22].
5	Tries to access to relations among main words by using the ontology [22].	Tries to expand the query by using keywords instead of using the methodology [25].

Table 2: Pros and Cons of Semantic Search Engines

No	SSE Name	Used Technology	Pros	Cons
1	Hakia	OntoSem (sense repository), QDEX (Query indexing method), Semantic Rank algorithm, Link & Free text	Determines information easily from trusted websites and gives accurate results	cannot index every individual, and needs helo from other search engines
2	Kngine	Uses fuzzy logic algorithm to determine the relations among concepts and the meaning of terms, synonyms, concept search	results appear as image in one form, it used many languages to allow the user to search in parallel. Categorized Results	The silent mode is not allowed and doesn't work in all browsers
3	DuckDuckGo	The results are found by either static data resources or 3rd party APIs. Clustered search, NLP	Does not record privacy of users' information and search for the results by many resources and web crawlers	Lacks video and image search
4	SenseBot	Uses text mining algorithm to identify semantic key concepts of searcher query and parses websites to perform a coherent summary	Summarizes different types of documents	Does not supports some browsers, only works with Firefox as extension, also display the results by Google
5	Swoogle	Indexing semantic, ontologies, OWL, RDF	Uses ontologies and documents for searching and provides metadata of SWDs	indexing of large data and queries are still challenge

3.11. Sensebot

The search process is to analyze web pages and to define the keywords depending on semantic concepts. It provides many of the documents that are related to search term and make summarization of the content of the documents to give the best answer to searcher [25]. The summary brings the best idea of some topic related to the searcher query. This summary is coherent to the searcher. Also, it saves time to provide the best topic related to search term and references to right resources. The engine tries to understand the whole concepts of sentence to give the suitable answer to searcher [18]. The searcher does not need to open many pages to meet his requirements [20].

3.12. Swoogle

Swoogle is an intelligent SSE that searches about the meaning of the words instead of the syntax. In addition, it is considered as a crawler depending on indexing and retrieval systems [24]. The structure of Swoogle is divided into four main components which are: 1) metadata creation, 2) Simple Web Discovery (SWD), 3) data analysis, and 4) the interface [27]. At the backend the SWD creates the database of the SW documents depending on a hybrid approach. It uses the address of URLs, URLs from conventional search, analysis SWDs and generates new URI candidates, this is used to generate URLs to find SWDs on the web. The indexing part is used to index SWDs depending on its metadata. The used techniques in this engine are RDF/XML, N-triple [16]. In addition, some other languages are also used such as OWL, DAML, PDFS and RDF. The analysis part of the engine used to create metadata to describe documents to classify the Semantic Web Ontologies (SWOs) and Semantic Web Databases (SWDBs) [19]. The last part of the system called services which considered as the engine interface that tries to provide data services depending on ontologies at the term level [27].

4. Literature Review

The SW is one of the important subjects that many of authors work within this area. These days everyone focuses on some features and techniques that uses the SSEs. Sahu et al. [12] made a comparison among four kinds of search engines according to their performance, they conclude Google as the best one has features. In addition, they give better results in most cases compared with other type of search engines because it uses the semantic query. Finally, they concluded that Google and Yahoo are developing every day but Bing is developed every month while Ask.com had become every old. The order from top to down: Google, Yahoo and Bing respectively.

Shah et al. [22] compared between different SSEs by using RDF technique in their approach. The RDF is depending on several classification criteria on SSEs. In addition, some technologies that used there are discussed with the evaluation of their performance. The advantage and disadvantage of each search engines are also presented there. The paper included analyzing the search engines' technologies and how a researcher can reduce the flaws for each engine. Finally, discussions for each engine as better, most suitable depending on the purpose, the quality of results in SSEs are presented and how do they need to improve day by day.

Malve and Chawan [18] concludes that the SSEs are better and have many advantages over the KSEs in terms of accuracy of presenting the results. The process of search in SSEs depends on semantic queries. In SW, the users have more assurance to achieve the accuracy of information and getting the answers based on the meaning of words that been searched by users instead the page rank algorithms and keywords. In addition, the main different between KSEs and the SSEs are presented. They also present the clear idea of the techniques used in SW which enables the user to achieve the best information.

Qureshi et al. [19] focus on exploring differences of SW search dimensions. They use the excellent pyramid to test the different dimensions to study about the SSEs. Even now the SSEs are in their developing stage and there are few numbers of resources in this field. In addition, there are many of explore search by querying for various device and difference semantic search engines record and stored formerly. All related materials used in semantic search can be obtained by many authors. Each of the search engines are depending on the pyramid of standard SSE. The authors also compare analysis between different of emerging search engines depending on pyramid which shows the requirements of the search engines.

Jain et al. [17] concluded that, the web 2.0 search engines are different from the SW search engines, because those in the web 2.0 search engines are unable to give the answers directly to user's query. The reasons of this problem are the web 2.0 search engines consist of unstructured information in nature while the web 3.0 (SW) uses RDF format to form the information in more suitable structure which helps semantic search engines such as Falcon, Swoogle, SWSE etc. understand the data and try to give more efficient results to the user. Web 3.0 deals with the data which is structured by RDF or OWL formats only. In addition, the web 2.0 also consists of large library of data linked together in semi-structured such as CSV and XML. The data in web 2.0 can rearrange in OWL format which can be benefited by the SSEs to expand the area of data search. Finally, the authors concluded that SSE technique is what is used in crawling, indexing, ranking and result formation process.

Jagtap et al. [20] briefly surveys many kinds of the SSEs which use different type of methods to search information for the user query. Furthermore, a comparison between the intelligent SSEs and their techniques, and the search engines which depending on high recall of perspectives but low of accuracy are presented. The determination of identify users,

Table 3: Comparison Based on Literature Review.

No	Refs.	Search Engines	Technologies	Purpose of paper	Conclusion
1	[12]	Google, Yahoo, Bing, Ask	OWL, RDF, XML, HTML.	Checks the speed of the question answer with the accuracy of information	Concluded that the Google is the best one that has features and gives a very quick answer
2	[22]	Intelligent Web Service Search, XploreProducts.com platform	RDF, SPARQL DBpedia	Explains the cons and pros, the best approaches selected of the SSEs	Compares among several SSEs according to their approaches and the criteria that used in each one
3	[18]	Wikipedia, Google, DuckDuckGo, Hakia, Swoogle	OWL, RDF, HTML, XML	Compares between KSEs & SSEs	The SSEs are better than KSEs and the results in SW is more accurate in results
4	[19]	Hakia, Lexxe, Cognition, Powerset, Sensebot	QDEX algorithm, NLP, RDF, RDQL	Compares among several SSEs according to various criteria such as Environment, Query Type, Intrinsic Problem	The SSE in the developing stage and they depending on the pyramid standard of SW levels
5	[17]	Falcon, SWSE, Swoogle, Hakia, Google	URIs, RDF, SWDs, OWL, XML, CSV	Explains the main features of several SSEs and the evaluation of each one	The search engines in Web 3.0 are better than those in the Web 2.0, and the SW is quicker to get the answer and the results related to user's request
6	[20]	MSN, Google, Proposed system, Yahoo	OWL, RDF, QDEXing, XSEarch, XCDSearch	Compares among several SSEs according to recall rate and precision rate	Concluded that the SSEs have different tools and techniques to search about the information with high recall rate and low accuracy
7	[21]	Google, DuckDuckGo, Kngine, Hakia	RDF, DAML+OIL, OIL and OWL, XML, SEWISE	To explain four approaches where the SSEs depend on, and the techniques that used of each one.	Concluded that some of SSEs use different approaches to decrease the exclusive of the use and to achieve to better information relatives to the user request
8	[22]	Swoogle, Factbites, DuckDuckGo, Sensebot, Kngine, Lexxe	RDF, OWL, QDEX (Query indexing method), fuzzy search, Semantic Rank algorithm	Presents a comparative study among techniques used in SSE	Searching within the internet is a challenge and needs new approaches to improve utilization of search engines

inaccuracy queries and crawler efficient, and the used tools in the SSEs for search of information on the websites are discussed as well. In addition, the development of SSEs is efficient and uses the technology to answer the complex queries of users. Also, the author makes short overview of the best SSEs which use different approaches in many methods to present the unique search experience for the users. The search on the internet today is a challenge because the most of the complex question unanswered while the SSEs present the suitable answers to user's queries.

Chitre in [21] presents some of the SSEs depending on different approaches using by different methods to decrease the exclusive search experience for the users. In addition, the search process on the internet today is challenging to predicate the efficient answers of the user queries that are suitable for the meaning. The author provides ways for how SSE can do better performance outperforms the limitations of the KSEs.

5. **Discussion**

In this section, a discussion for the reviewed search engines will be presented. As shown from table 1, the main differences between SSE and KSE is that, KSE is based mostly on conventional technologies such as HTML, XML... etc. while SSEs are using Semantic Web technologies such as RDF, OWL…etc. As given in table, all of the systems use RDF, OWL and other semantic web technologies. As it is obvious from Table 2, the mutual advantage of the SSE is that all of them retrieve accurate results to the query, while the common disadvantage is that indexing of large chunks of data is a challenge. KSEs are simple to implement and most of the users know how to use it, while SSEs are more complex systems and need the technology of Semantic Web to be implemented. As shown from table 3, most of the SSEs are not yet common to the public, this is due to a good knowledge is required for lay-users to let them use it which makes it complex for them. The giant search engine companies such as Google and Bing started moderately to include Semantic Web in its search results so that the SW concept becomes more familiar for normal users.

6. **Conclusion**

The current Web offers an easy way to share information online, this makes the size of data on the web become huge gradually. Search engines help users to find information on the web. There are two types of search engines: Keywords Search Engines (KSE) and Semantic Search Engines (SSE). The KSE considers as the base search engines of the web, but they cannot find exact and accurate information to the user queries because they depend on the syntax of the keywords. SSEs solve this problem by looking into the meaning of the keywords and retrieves related results semantically. SSE depend on the technology of the Semantic Web which help them understand the concepts and help machines to

understand and process information. In this paper, an overview of Web technology is given, search engines and their types are presented, then a comprehensive comparison for the two most common search engines KSE and SSE are exposed. Finally, a wide-ranging discussion for each of the reviewed systems with their used technologies, techniques, methods, pros and cons are presented in details.

References

[1] K. Jacksi, S.M. Abass, "Development History of the World Wide Web."

[2] A. Gyrard, M. Serrano, G.A. Atemezing, "Semantic web methodologies, best practices and ontology engineering applied to Internet of Things," IEEE: 412–417, 2015.

[3] K. Jacksi, S.R.M. Zeebaree, N. Dimililer, "LOD Explorer: Presenting the Web of Data," International Journal of Advanced Computer Science and Applications (IJACSA), **9**(1), 2018, doi:10.14569/IJACSA.2018.090107.

[4] K. Jacksi, N. Dimililer, S.R. Zeebaree, "State of the Art Exploration Systems for Linked Data: A Review," International Journal of Advanced Computer Science and Applications (IJACSA), **7**(11), 155–164, 2016, doi:dx.doi.org/10.14569/IJACSA.2016.071120.

[5] A. AL MUQRISHI, A. Sayed, M. KAYED, "CASENG: ARABIC SEMANTIC SEARCH ENGINE.," Journal of Theoretical & Applied Information Technology, **75**(2), 2015.

[6] S.R. Zeebaree, A.-Z. Adel, K. Jacksi, A. Selamat, "Designing an ontology of E-learning system for duhok polytechnic university using protégé OWL tool," Journal of Advanced Research in Dynamic and Control Systems, **11**(05–Special Issue), 24–37, 2019.

[7] K. Jacksi, N. Dimililer, S.R.M. Zeebaree, "A Survey of Exploratory Search Systems Based on LOD Resources," in: , in PROCEEDINGS OF THE 5TH INTERNATIONAL CONFERENCE ON COMPUTING & INFORMATICS, UNIV UTARI MALAYSIA-UUM, COLL ARTS & SCI, INFOR TECHNOL BLDG, SINTOK, KEDAH 06010, MALAYSIA: 501–509, 2015.

[8] A. AL-Zebari, S.R.M. Zeebaree, K. Jacksi, A. Selamat, "ELMS–DPU Ontology Visualization with Protégé VOWL and Web VOWL," Journal of Advanced Research in Dynamic and Control Systems, **Volume 11**(01–Special Issue), 478–485, 2019.

[9] D. Fensel, "Ontology-based knowledge management," Computer, **35**(11), 56–59, 2002.

[10] A. Akerman, J. Tyree, "Using ontology to support development of software architectures," IBM Systems Journal, **45**(4), 813–825, 2006.

[11] J. Rashid, M.W. Nisar, "A Study on Semantic Searching, Semantic Search Engines and Technologies Used for Semantic Search Engines," International Journal of Information Technology and Computer Science (IJITCS), **8**(10), 82–89, 2016.

[12] S.K. Sahu, D. Mahapatra, R. Balabantaray, "Comparative Study of Search Engines In Context Of Features And Semantics," Journal of Theoretical and Applied Information Technology, **88**(2), 210–218, 2016.

[13] Z.N. Rashid, S.R. Zebari, K.H. Sharif, K. Jacksi, "Distributed Cloud Computing and Distributed Parallel Computing: A Review," IEEE: 167–172, 2018.

[14] R. R. Zebari, S. R. M. Zeebaree, K. Jacksi, "Impact Analysis of HTTP and SYN Flood DDoS Attacks on Apache 2 and IIS 10.0 Web Servers," 156–161, 2018, doi:10.1109/ICOASE.2018.8548783.

[15] R.R. Zebari, S.R. Zeebaree, K. Jacksi, H.M. Shukur, "E-Business Requirements For Flexibility And Implementation Enterprise System: A Review."

[16] S.C. Damle, S.S. Kupekar, K. Deulkar, "Comparative Study of Semantic Search Techniques using RDF," International Journal of Computer Applications, **975**, 8887, 2015.

[17] R. Jain, N. Duhan, A. Sharma, "Comparative study on semantic search engines," International Journal of Computer Applications, **975**, 8887, 2015.

[18] A. Malve, P. Chawan, "A Comparative Study of Keyword and Semantic based Search Engine," International Journal of Innovative Research in Science, Engineering and Technology, **4**(11), 11156–11161, 2015.

[19] M.M. Qureshi, B. Asma, H.U. Khan, "Comparative analysis of semantic search engines based on requirement space pyramid," International Journal of Future Computer and Communication, **2**(6), 562, 2013.

[20] D. Jagtap, N. Argade, S. Date, S. Hole, M. Salunke, "Implementation of Intelligent Semantic Web Search Engine."

[21] N. Chitre, "Semantic Web Search Engine," International Journal Of Advance Research In Computer Science And Management Studies, **4**(7), 2016.

[22] V. Shah, A. Shah, K. Deulkar, "Comparative Study Of Semantic Search Engines," International Journal Of Engineering And Computer Science, **4**(11), 14969–14972, 2015.

[23] S.R. Ranganathan, M. Marrikkannan, S. Karthik, "NEED OF FUTURE WEB TECHNOLOGY THE SEMANTIC WEB A BRIEF SURVEY," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), **1**(10), pp-71, 2012.

[24] M. Rahman, "Search engines going beyond keyword search: a survey," Int. J. Comput. Appl, **75**(17), 1–8, 2013.

[25] S.S. Laddha, P.M. Jawandhiya, "Semantic search engine," Indian J. Sci. Technol, **10**(23), 01–06, 2017.

[26] G. Sudeepthi, G. Anuradha, M.S.P. Babu, "A survey on semantic web search engine," International Journal of Computer Science Issues (IJCSI), **9**(2), 241, 2012.

[27] S.K. Sahu, N. Kumar, S. Allawadhi, J. Singh, "Comparative Analysis of Semantic and Syntactic Based Search Engines," Journal of Advanced Computing and Communication Technologies, **4**(2), 22–27, 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 112-118 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Internalising Negative Self-Image Externalities: The First Objective for City Marketing as a Municipal Management Tool[1]

Manuel Escourido Calvo*, Antonio Javier Prado Domínguez, Valentín Alejandro Martínez Fernández

Faculty of Economics and Business, A Coruña University, 15008, Spain

ARTICLE INFO

Article history:
Received: 26 November, 2019
Accepted: 04 January, 2020
Online: 15 January, 2020

Keywords:
city marketing
externality
self-image
sense/pride of belonging
perception

ABSTRACT

In times of crisis or traumatic transformation processes, one of the most frequent negative externalities is that the self-image of a city deteriorates among its stakeholders, which affects their economic and social expectations and decisions. The internalisation of this externality must be a key objective for the local public manager so that any initiative can be carried out with collective support and collaboration. When applied to this end, city marketing and the instruments that this technology makes available to city managers provide optimal results.

1. Introduction

From a historical perspective, cities are understood as the domains of human interaction in which civilisation develops; for this reason, they play a role in supporting economic activity [1]. In fact, an economic analysis of civilisation must take into account its concreteness: that is to say, the city itself. This is the reason why many disciplines approached the subject through the study of the city and its management, knowing full well "the city is, manifestly, something very complex" [2].

The importance of "good" city management is manifest, above all, in times of economic crisis. The crisis of 1973 led to the adoption of a new conceptual framework for city management during a time of restructuration, with a need for social policies aimed at job creation, urban renewal policies and economic policies focused on externalities. Years later, the 1990's industrial depression meant that the city had to make city management decisions based on new technologies, innovation and the negative impact externalities had on urban ecosystems.

Urban managers began to acknowledge the importance of remaining competitive in the city global market during the 2008 crisis under the threat of population decline, unemployment or the loss of investment in favour of more competitive cities.

The city is the crossroad where interests and actions meet and where economic, social, urban and environmental conflicts are expressed with great intensity. When this scenario depends on the fate of industry or industrial groups and resources and/or local produce is in decline- in terms of job creation and invigoration or economic engine- the challenge is even greater.

This has happened in many cities that went through and/or have ongoing industrial restructuring or termination/closure procedures related to the natural resources on which they were dependent. In turn their downfall triggers a demographic threat, with the loss of population in the age bracket that has the biggest impact on birth rates. Decline can also be expressed in economic terms with the loss of future, high quality taxable base income, added value workers, entrepreneurs and externalities, as well as in social terms, with an overall loss of talent.

As explained in research by Escourido, M. [3], which is part of his PhD thesis, this environmental change and foundational transformation of the city resulted in an enormous amount of public funding being spent on discretionary actions in an effort to solve a multifaceted problem with only one solution. The problem was coupled with the absence of a systematic and competitive diagnosis of the city and a much needed strategic thinking process focused on both supply and demand. These actions were detached from any planning and had no impact on the city, although political marketing slightly benefited by them.

*Manuel Escourido Calvo, Email: manuel.escourido@udc.es
[1]This paper is an extension of work originally presented in the 2019 14th Iberian Conference on Information Systems and Technologies (CISTI)

The process of change facing the city this century happen in such a quick and constant pace that, in many cases, the community is overwhelmed and unable to react. If there are ongoing, dramatic situations in the city, the challenge is greater. Nevertheless, cities in decline and lacking in expectations also have citizens asking for solutions to problems related to employment, education, safety and social services. Their demands reach urban managers. It is these public servants who must allocate suitable services to meet their needs and improve the ranking of their city. Being unsuccessful in meeting these goals means that locals can "vote with their feet" [4].

We must not forget that cities are made up of the people who define them, adding dimensions and guiding them to their future. This is why problematic areas are not only those with business or industry in decline, but also those where people think the future will be the same as today [5].

We are then looking at city management with certain features. Its social and economic development process originated from public sector activity, such as in a state energy policy and/or industrialization. Moreover, the economic industrial monoculture model drifted because the market and the comparative competitiveness of factors and sectors went into decline. Even if this trend had been caused by market failure, it suggest that the public sector had erred in its passivity [6].

On the other hand, interventionist measures are overlooked or left incomplete when the city is faced with a socioeconomic change of such magnitude that it leads to a series of negative consequences. Among these outcomes are a highly negative externality and an absence of expectations. Another is the negative self-image of the city, which mirrors its dwindling attractiveness and its social and economic decay. These problems can be offset by changing how the city is perceived with marketing based on urban management tools funded with public-private collaboration formulae.

2. Coase and the externality internalization (solution)

State intervention through regulations, bans and taxation is the usual response to the externalities problem [7], always under the wrong However, in doing so, the public sector determines legal responsibilities instead of clarifying the economic nature of the problem. Thus, the system's efficiency may be diminished.

This happens because the State and courts argue over who has the right to take action rather than decide what should be done. These initial legal rights boundaries can always be modified through market transactions, redefining themselves in the process if this leads to increased production.

At this point, one can follow the thinking of Coase. He refers to, but questions, the assumption that there are pre-existing property rights, assigned and defined in a clear manner, with inexistent or low transaction costs. Enshrined in law, these property rights are in conflict with a viable and efficient economic solution. Thus they would be reallocated towards those who value them the most, even if the legal system goes against them.

With zero transaction costs, economic agents would elaborate the contractual ruling necessary to maximise production value, whatever the nature of these circumstances may be: legal, with liabilities or through wealth distribution. At the same time, Coase [8] states that, when real costs are generated and must be accounted for, many of these agreements and deals may be annulled because their cost would be higher than the income generated.

Taking into account transaction costs, Coase points out that both parties may feel their incentives are lessened when the information needed to carry out deals and agreements have been revealed. Public intervention would be needed to define which incentives are missing or which contractual deals can be done. Even in the case of exceedingly high transaction costs, state intervention can protect specific activities that are causing harmful externalities.

In this context, a point may have been overlooked: communal private property- mainly business- is defined as a system aimed at minimizing transaction costs, even more so in complex social environments.

In conclusion, this is a general balance model for a market economy, whose theoretical underpinnings are closer to being defined. They aim at maximizing efficiency, without taking into account the reason or use of property rights. The State still has the right to act when transactions have high cost or pose difficulties. Otherwise, these agreements will be untenable. Another advantage of the Coase Theorem is that the aforementioned situation is suitable for the externalities of production or consumption. It also comes into play in cases of inaction, when a forest has not been maintained properly or if a government fails to foresee the socioeconomic consequences of a declining industrial monoculture.

Nevertheless, Coase warns that the proposition is weakened by the sense of instability in the balance; the economy is defined as a process, not a state.

With these arguments in mind and heeding Cooter and Ulen [9], one may reach the conclusion that the State's main function is be to minimise the cost of every economic transaction, wherever it may be. In most cases, property rights would be clearly defined. This principle, known as Coase's Normative Theorem, means that law is structured "in such a way that nothing prevents private agreements from being reached". Apart from facilitating and promoting deals, it should minimise losses caused by the lack of deals [10].

3. Including the particularities of subjectivism and perception in internalising (solving) externalities.

Some of the objections to Coase's Theorem are related to the concept of human preferences. These objections, like human preferences, can be subjective, dynamic, shifting and therefore not measurable or comparable. Every market agent would try to increase their "psychological profits" in search of solutions that make them possible. This is mostly a problem of information and perception, as well as a strategic game and aversion to risk. Even taking zero transaction costs into account and properly defining private property rights, the potential possibilities of each deal would have to be taken into account to solve an externality.

As stated by Hayek [11], social science measures human perception around the world in terms of people's beliefs. A true or false perspective cannot be observed directly; the speech and

actions of an individual, however, can be studied. Human actions are not an objective fact: they are what the person performing them believes those action do, and so they form part of a personal, individual perception. Therefore, the perception of goals and means is introduced as an element to consider when solving externalities.

The same applies to cost and risk perception because the elements of individual choice depend on new ideas and are taken into consideration by the person who decides them and his or her preferences [12]. It is worth noting that, in a current behavioral analysis of Coase's Theorem, one of the results, among others, suggest that, when an "unfair offer" is presented, the desire to break negotiations is so strong that the Theorem does not apply, even when it is against the self-interests of the individual rejecting the offer [13].

It is therefore possible to ascertain that social cost and social products are not objective concepts; different individuals consider benefits and prejudices based on their own perceptions, presumptions and analysis. This means efficiency in the choices made cannot be assessed in an objective way.

At this point, the market has tools- including profit and competition as motives- which provide the incentives needed to define and then meet the needs and desires of consumers. The successful provider earns profits, supplanting the others who fail at this task. These tools do not guarantee that the correct choices are always made, because they cannot be known. However, in the long run, it is expected that these same tools optimise the social net benefit [14].

Public sector management must step in at this point so that it can identify, prioritise, and meet the needs and demands of clients or those managed. To this end, marketing as an urban management tool should be adopted. With the paradigm shift ushered in by New Public Management, the administration focuses on the "client-citizen". The end result is custom-made solutions for citizens, not only in the domain of goods and services but also in the promotion and creation of incentives, particular conducts and desirable perceptions, such as those contributing to the internalisation of specific externalities.

4. Externalities of negative self-image. As Pontes de García Rodríguez (Spain) as a case in point

To favour industrialisation and empower energy production, state economic intervention during the 1970's in Spain favoured lignite mining in As Pontes de García Rodríguez (A Coruña, Galicia, NW Spain). This was done in an open pit mining operation. The mineral was then used as fuel in a power plant that was built by the public sector in the same city. A public company was responsible for the project: Endesa (Empresa Nacional de Electricidad SA). In what had once been a rural city, a period of demographic and economic expansion began. It had a striking impact on the income and wealth of the area.

On the one hand, expropriations were carried out to establish the mining operation, which provided unimaginable wealth to mostly rural land owners with forests, crops and livestock. In many cases, this change also meant access to what resembled public employee posts in the mines, power plant or the services attached to them. Moreover, salaries were well above the average national wage, with complements and benefits: the state paid for accommodation, schooling for employees' children and workers' personal electricity bills. In the same fashion, the income generated by this company in the public treasury meant the council budget continued to grow. As a result, an infrastructure was created to support services whose quality was higher compared to that in other cities of similar size and population.

However, all the mining operations and power plant activity overshadowed local businesses due to a difference in employment opportunities and salaries offered by Endesa. A great amount of resources was allocated towards this company. The primary sector, in contrast, was practically reduced to subsistence farming of crops and livestock in those places where the land was not expropriated by the state for mining. On the other hand, the construction sector developed small companies driven by the high demand for housing for the newly arrived workers for the mines and power plant.

Nevertheless, this development was limited by Endesa itself as it was a public company back then. Endesa built over a thousand homes for employees and their families in exchange for a symbolic rent. At the same time, an auxiliary and service industry for repairs, maintenance and supplies emerged, linked to and dependant on mining and energy production. This sector had a presence in commerce, thanks to many catering services.

This situation of extreme economic dynamism in the municipality and surrounding areas and the positive externality resulting from public sector intervention made a complete turnaround after two events. Firstly, the mining operation was closed in December 2007 in accordance with European Union directives. The second event is the worldwide crisis, as experienced in As Pontes de García Rodríguez, especially from 2012 onwards. The shift in focus of environmental legislation in the European Union conveys a negative externality leading to the direct closure of the mining operation in this municipality. Even more, it was known that lignite exploitation would be curtailed and public intervention was limited or non-existent so little effort was made to minimise the risks arising from a social and economic industrial monoculture with a limited lifespan.

Since then the city has felt the effects of socioeconomic decline: unemployment, retirement and early retirement from the mining operation, a lack of opportunities for the younger population brackets with higher education and the failure of re-industrialisation initiatives despite having the advantages of heavy public funding. Moreover, in the absence of planning and management, the different public administrations have been unable to turn the tide in a place that was "killed by its own success".

In turn, the city faces a demographic threat with a loss of population in the age bracket that has the biggest impact on birth rates; an economic threat with loss of future high quality taxable base income, added value workers, entrepreneurs and externalities and also a social one with an overall loss of talent. These threat were paired with the disappearance of retired workers and their retirement and early retirement pensions. This led to the risk of an urban decay process: reduced population, fewer services, more taxes and a less stable population.

As Pontes de García Rodríguez represents many of the points mentioned earlier:

- Here, city management focused on public-financed social and economic development from the public sector leading to positive externalities.

- However, in the absence of a collective vision for the city, public management failed to exploit multiple synergies while they were still viable to break away from the economic model of a monocultural industrial reality.

- There were clear indications of the public sector's passive attitude, whose intervention was non-existent, incomplete or misguided when challenged with a socioeconomic process of decline on such a large scale.

- Its most striking externalities (at least in regards to the priority that its internalisation must have) demonstrate a lack of expectations and negative self-image of their local population.

In fact, research was carried out by Escourido [3] through telephone conversations in a poll with a sample of 9,451 residents over 16 years of age, with a confidence level of 95% and 5% maximum sample error margin. Also interviewed was a panel of city representatives (Delphi, with total of 21 individuals). This study reveals low expectations with regards to the future of the city, based on the perception of the citizens.

This is considered a city with an ongoing deindustrialisation phase and a decadent image. Moreover there was a sensation of improvisation and inability from the public sector, as well a lack of social cohesion brought on by the duality between retired/early retired miners with high pensions and the rest of the population, affected by the loss of employment.

The internalisation (solution) of an externality closely related to the perceptions, desires and needs of the individuals has to take into account the need for public intervention aimed at reducing or eliminating transaction costs (information, property rights reallocation and negotiation). This must be carried out with the use of public and private collaboration mechanisms, marketing technology tools for cities which boost its self-image and strengthen a sense of belonging to the city. All of these measures are essential to develop subsequent socioeconomic recovery initiatives.

5. City marketing as a tool for public management to internalise the externalities of negative self-image. Case study of As Pontes de García Rodríguez (A Coruña).

Traditionally public management appeals to the political process for resource allocation; criteria include redistribution, equity, the common good or general interest. The level of focus depends on plurality, as well as administrative and executive fragmentation. Allocation, therefore, comes from legal actions and processes. In value creation, many factors are taken into account. These are variety and attachment, as well as transparency, zero discrimination, legality and social perceptions. This is especially true when the administered citizen acknowledges that the collective interest is not the sum total of all the individual interest, contrary to the way it is in the market when the word "client" is mentioned.

With economic efficiency goals -such as continuous improvement, resource utilisation and process based organised structure- and the citizen-client service, public management is more than a sovereign, coercive instrument. Instead, it has the end goal of providing services for society, such as identifying citizen needs and meeting them in an efficient and successful manner and with the highest possible quality. Marketing then becomes a technology that is adopted from the private sector for the public administration, to work as a detector and a way to measure the needs of the target market. Services can thus be developed to meet their needs and prepare for the upcoming changes. Every step is done in an efficient and successful way under the guidance of its vision, mission and values and goals.

This is the major advantage of city marketing when it is compared to other planning tools. The traditional focus of supply is supported by one of demand. This mechanism can spot and solve current needs by providing adequate services and infrastructure and future ones as well by developing new assets to improve living standards and keep competitive advantages in the target markets and establish communication in a way that a city brand is built. Moreover, the focus on demand allows for a more realistic and less resource intensive framework of action, with advantages derived from its efficiency.

One of the goals of city marketing is to improve and develop a positive city image [15]. Image has an enormous impact on city life as a "determining factor in urban development" [16]. This can become either a huge supporting factor or obstacle for the economic, cultural and social development of a city. It is considered a high importance "asset" in the development of city management [17].

This city image is built upon when defining its "vision". This is a must when planning projects based on demand (target market) and on identity resources. These are ideal for creating a guaranteed difference in the product [18]. The results will depend on the accuracy of the identity perception and the suitability of how the demand is adapted to each case. It is important to know beforehand the image of the city from the perspective of its locals and users: self-image is the starting point. Once the self-image is known, differences between the city's goals and those of its citizens, business and institutions can be asserted.

As mentioned above, if the starting point is a negative self-image of the city-no matter the origin of these externalities- the top priority should be to modify and restore the self-image in a positive way. This is also to restore a sense of belonging, along with local pride, to the city. It is the only way to make citizens take part in restoring the city and bring stakeholders back to plan for its future vision. With city marketing, the first objective is to define the city project based on the demand for its identified resources. This is the reason why a positive internal perception is needed in the city to minimise risks while elaborating future plans.

It is true that certain tangible elements of the city are required, like infrastructure or services, but these are not enough anymore. Nowadays intangible resources are what adds value and makes a

difference [19] when planning how to manage its capacity to attract people, investors, business, administration and tourism. In this sense, the top priority must be to fortify citizen-city relationships. The opinion of satisfied citizens about the city where they live creates a positive self-image of that city. Making sure every citizen has a certain living standard and services that meet their expectations and desires transforms them into active communicators of the city when they are outside of it. That is why it is vital to analyse the self-image of the city, to reinforce it or to correct a possible negative value.

When image policies are supported in a realistic vision of the city and the active participation of all its citizens is promoted by the administration, a brand image is created in a stronger, more efficient fashion. This is the opposite of a brand created solely by the local public servants. Image policy must strike a balance between image and quality of life, jointly coordinated by those responsible for that image policy, citizens and the municipal government.

Management of the image falls into the hands of its inhabitants, who are the inside target market, a vital task with the goal of fixing, improving and restoring a city's self-image. If the different inside target markets of the city are satisfied with or proud of the city they live in, they in turn help to make a better image and positive brand. If living standards, along with the social, cultural and economic environment, meet the diverse needs of citizens, business, and local institutions, each one of them will become a definer- or in marketing terms, a "preacher"- of the city brand.

To achieve this, internal communication must be suitably systematised so that information flows and positive effects are generated among all citizens and institutions within the city. This strengthens the values associated with pride and sense of belonging to or living in the city, innovation and change. Also, a positive self-image is the lever needed to involve, motivate and engage the stakeholders of the city with the city project. In the best case scenario, this project will be the product of participation, general agreement and internalisation.

This is a bottom-up approach, where city marketing expands the possibilities of strategic management and provides local authorities their participation in the hearings, discussions and decision making those groups and sections of the population who detect and pose in front of city problems [20], starting with the solution for the negative self-image.

In the case of As Pontes de García Rodríguez, there is access to different tools and resources to accomplish this internal communication. The city also has to confront one of the greatest challenges during the research phase. In particular, the panel of city representatives has to counter the existing negative connotations of the city's self-image. What the city is for its internal target markets and what they would like the city to be is an aspiration that must be present at all times. At the planning stage of the strategic city marketing, this must be true. These ideas must also inform the operational stage, when deciding which actions should be taken to meet the needs detected in the target markets of the city.

When the City Marketing Plan and all the measures mentioned in it are implemented, a general agreement must exist, with all public and private officers involved. This guarantees the continuity of the plan far beyond electoral horizons. The marketing applied to city planning means a long-term strategy aimed at creating new competitive advantages, one of which is the city brand itself. Also needed is the general support of all public and private agents within the city. To accomplish this, the contents of this plan must be agreed on by everyone and widely disseminated. Every strategic option planned for the city must have the approval of its stakeholders to avoid their rejection and with it the failure of its implementation.

The city and its planning are not described in terms of economic productivity and capacities. Instead, the focus should be on its attractiveness and the city's position in different global networks; this approach requires considerable collaborative and creative input. With the use of city marketing as a tool for local public management, interest (or what it is expressed as value) is the potential on a micro scale (the city) and as such it can be adapted to a macro scale (international flow system) to achieve a position as an attractive and well placed city in the global network. To do this, the city brand and its self-image and territorial coherence are essential, as is the sense of belonging, identity and planning to achieve it.

All without forgetting that the image of the city can be defined as organic, in continuous change [21]. If this image is negative (for example, after a process of long industrial decline), it requires a comprehensive application of city marketing (and not just advertising). This is because it is necessary to modify the self-image of the inhabitants to then be able to involve them in the transformation process that the city needs.

In this particular case (described in point 4), to improve self-image as well as recover self-esteem and sense of belonging among citizens, there is a series of measures that are to be carried out using city marketing as a tool for local public management [22] wich provides solutions for negative externalities. To correct these externalities, steps include [3]:

a) *Building a city network.* The objective is to create a configuration of networks allowing pressure groups to claim leadership positions and promote cities in a broad and ambitious way. It serves as a tool for implementing advanced technologies to gain access to a larger volume of information and exchange experiences at every city management level about supply and services, government, efficiency and management quality, environment and other topics. This involves integrating into and being part of an urban economic type network to find solutions to common problems and identify competitive advantages or share urban innovations. With cultural concerns, there can be promotion and defence of joint interests. The tourist domain is to promote tourism, culture, history and the environment, while the social content is about promoting values and cooperation. This is a communication and management tool for external projection, making the city more attractive and, above all, improving its self-image. The associations around an exemplary city and its attributes improve its perceived image both in- and outside that place.

b) *Promoting local heritage.* The historic, natural and industrial heritage in As Pontes de García Rodríguez is an element that connects past, present and future and projects the city's identity as a collective and/or a city. The footprints left behind by history, human interaction with nature and intense industrial activity in the area are a common link that can be used as a powerful tool to educate, build awareness and communicate in terms of an identity, an image and/or a brand. This natural and industrial heritage is used to preserve identity and self-image, while maintaining the continuity of the community and, at the same time, generating the right to use it and the collective duty of preserving it for the future.

A plan must be put in place to emphasise the value of its historic heritage, with archaeological sites from the Palaeolithic and Neolithic age, Roman period objects and bridge and the noble patronage of García Rodríguez de Valcárcel; its natural heritage, with is natural park in the Fragas do Eume region, lake, coal heaps, A Ribeira and Eume reservoirs and Caxado hills; and industrial heritage linked to mining operations and their industrial development. A process is needed to select what is of value and determine the next steps in incorporating collective identity and the actions that have to be taken to preserve and promote it.

It is important not to lose sight of the interpretation and value of these elements. This is done through logic, interests and perceptions because the planning must take into account general agreements about concepts, management models and value-deposed formulae.

c) *Reclaiming talent*. Building awareness to raise the level of talent inside the city lets it improve its self-image, as well as develop a sense of belonging and a collective pride. To shift perceptions and make the city more attractive to locals, events- such as conventions, shows and meetings- can be held to gather local people who excel professionally and have management positions in business or organisations.

d) *Mobilise and bond with the migrants*. As Pontes de García Rodríguez lived through the same migrations as the rest of Galicia during the end of the 19th century and beginning of the 20th (mostly migrants that went to South America) and again during the 50s and 60s (to South America and Centre of Europe). Later, the mining operations and power station made by Endesa, combined with the economic boom, allowed the vast majority of young people to have access to training and education at all levels showcasing a high percentage of citizens that reached higher education. Most of those students failed to find a job in their local city so they went to other areas.

At this point, the goal is to facilitate communication channels between "ponteses" (As Pontes locals) who are outside the city and keep and promote ties with the city. This measure can become an ideal economic, social and cultural tool among the members of the city and the institutions. These goals help to develop the city identity of the locals and their environment (family, social and economic); outside perception improves as the self-image of the locals gets better.

As shown, the courses of action proposed are aimed at public and private interests alike, during a time when people are solving public problems based on perceptions and personal feelings. This entails taking measures and holding discourse that brings satisfaction and meets needs. Also needed are a public worldview and narrative discourses (even political ones) that go deeper into the construction of the city from the public sector perspective.

With all of these points seen as tools for city marketing, the city is a space and an ensemble of meanings that are shaped and given function and purpose as the social agents relate to each other. They create a space to understand the diverse ways to express oneself, interpret symbols and produce meaning. This is where every audience is built from the collective senses, cooperation towards social change and encouragement of positive self-image and city pride.

Our perception of the world is almost always local, focusing, exclusively, on what is around us and our neighbourhoods, traditions and beliefs. At the same time, aspirations exist in all of our levels of society, from the lowest to the highest. Every one of us sends signals that express belonging to a tribe. If the big brands appeals to perceptions, needs, desires and feelings or hope, attractiveness, desirability, love, acceptance, and innovation, the image and brand of a city does not differ. As an individual and also as a citizen, I long for "what we had and lost" and "What we used to dream/be and never had/were" [23]. How we value this longing, evaluating and identifying the city and the image we have of it. To sum up, every community craves a strong sense of belonging and pride, which is the vehicle to any process of social change. This shows the importance of management from the local public perspective, the self-image of the city.

The aphorism "whatever the locals find unattractive outsiders will as well" is especially true in the case of As Pontes de García Rodríguez, as shown in the cited research. The appeal to city marketing from a public management perspective makes it possible, among other things, to act upon the immaterial values or intangibles of the city, in such a way that the self-image of the city shifts into something attractive for the residents. And it makes it possible, even in its application to small and medium-sized cities [24].

To do this, improving the reputation and citizen self-esteem, achieving a welcoming social environment open to innovation and lower the resistance to change (or adaptation) are the first steps needed to lay the foundations. Thus the main goal of public marketing is reached: enhancing the attractiveness of the city.

6. Conclussions

The importance of "good" management of the city is manifest, especially during a crisis. Traditional strategic planning focused on supply is not enough. An approach that focuses on demand is needed, along with the use of a city marketing technology.

In cities where the last recession has generated traumatic processes in terms of restructuring or the social and/or economic fabric has disappeared especially when they once enjoyed times of intense economic growth, one of the negative externalities with highest impact is the emergence of a negative self-image in the city. This in turns affects the city brand, expectations and decisions of its stakeholders and, with them, the future development of the city itself.

These externalities come from market failures and/or public sector mistakes and require public intervention to minimise or eliminate transaction costs, as Coase pointed out. Taking into account the nuances of subjectivism and perception management helps internalise (solve) the problem. This should be done by the citizens, and from there, act as a lever to promote social change and the future development of the city. To do this, city marketing used as a public urban management tool can be a solution for these externalities. When it is present, this must be done before implementing the rest of the City Marketing Plan, because the restoration of a positive self-image, pride and sense of belonging is a requirement to mobilise, integrate, involve and engage the stakeholders of the city to help the city reach the "vision" goal via the implementation of the rest of the plan

Every element is present in the case of As Pontes de García Rodríguez (Galicia, NW Spain). Public intervention made an industrial monoculture model that generated wealth and jobs for decades. When this model was exhausted and the state failed to intervene on a suitable scale, the city reached its present situation of decline. This decline is socioeconomic in nature, where the main externality is the generation of a negative self-image of the city, leading to a lack of expectations, depopulation and distrust in public management.

The use of city marketing as a tool for local and public management proposes as the top priority those actions that interiorise (solve) the negative self-image externalities in the city as the first step before anything else is done. This precedes any social or economic change in the city, and has to be done and planned for, taking into account the general approval and participation of the citizens who believe in the future and a vision for their city. This is a bottom-up approach.

It is important to note that marketing and the city are both in a continuous process of change and evolution. For this reason, the diagnosis stage could be carried out at a deeper level and report current perceptions of self-image through the use of new technologies: big/small data, the internet of things or social media.

At the same time, it is necessary to assess in which ways the immaterial aspects of the city gain importance over the material and to see how they flow, influencing the definition of new shapes and forms of city planning and management. This can be done by integrating new methodologies of citizenship participation and social creativity. It would be useful to contrast "city branding" from the point of view of public policy supported by the strategic planning of the city to enhance its capacities and unfold its competitive advantages.

References

[1] R. Camagni, Economia Urbana, Antoni Bosch Editor, 2005.
[2] D. Harvey, The Urban Experience, Basil Blackwell, 1989.
[3] M. Escourido, "El marketing de ciudades como herramienta de gestión pública local: una aplicación al caso de As Pontes de García Rodríguez (A Coruña)", Ph.D Thesis, A Coruña University, 2017.
[4] C. Tiebout, "A Pure Theory of Local Expediture", The Journal of Political Economy, 64 (5), 416-424, 1956.
[5] P. Kotler, D. Haider, I. Rein, Mercadotecnia de localidades, Ed. Diana, 1994.
[6] D. Weimer, A. Vining, Policy Analysis and Concepts, Prentice Hall, 2004.
[7] A. C. Pigou, The Economics of Welfare, MacMillan, 1932.
[8] R. H. Coase, La empresa, el mercado y la ley, Alianza Ed., 1994.
[9] R. Cooter, T. Ulen, Law and Economics, Harper Collins, 1988.
[10] J. M. Aguilera, "El análisis económico del derecho de propiedad: una aproximación", Anuario Jurídico y Económico Escurialense, XLII, 117-128, 2009.
[11] F. A. Hayek, The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology, Kessinger Legacy Reprints, 1952.
[12] L. M. Kirzner, "Competition and Entrepreneurship", University of Chicago Press, 1973.
[13] R. Thaler, Todo lo que he aprendido con la psicología económica: el encuentro entre la economía y la psicología, y sus implicaciones para los individuos, Ed. Deusto, 2016.
[14] S. C. Littlechild, "El problema del costo social", Revista Libertas 12, Instituto Universitario Eseade, 1990.
[15] M. Escourido, "La marca ciudad: la importancia de su valor", Revista ACEF-UDIMA, 2013.
[16] R. Friedmann, "Marketing urbano", Contactar, 3, 1999.
[17] M. A. Akbar et al., "An empirical Study Investigation of Task Allocation Process Barriers in the Context of Offshore Software Development Outsourcing: An Organization Size Based Analysis," Int. J. Comput. Digit. Syst., vol. 8, no. 04, pp. 343–350, 2019.
[18] A. Precedo, M. Escourido, Marketing de ciudades y territorios: una visión renovada, Dextra Ed., 2017.
[19] S. Rivero, "Importancia de los intangibles como motores del crecimiento metropolitano", Curso Gestión Estratégica de Ciudades, Asociación Bilbao Metrópoli 30, 2001.
[20] A. Kiseleva, "Marketing Strategy of the Territory in the Investment Policy of the City", Strategic Management, 22 (1), 22-29, 2017.
[21] G. Hankinson, "Managing Destination Brands: establishing a Theoretical Foundation", Journal of Marketing Management, 25 (1-2), 97-115, 2009.
[22] A. Martin, "Marketing a Municipality as a Place. A Study of Three Municipalities in the north of Sweden", Bachelor Tesis, Lulea University of Technology, 2009.
[23] M. Lindstrom, Small Data, Ed. Deusto, 2016.
[24] N. Dril, A. Galkin, N. Bibik, "Applying City Marketing as a Tool to support Sustainable Development is Samal Cities: case estudy in Ukraine", 2nd International Conference "Green Cities – Green Logistics for Greener Cities", Szczecin, Poland, 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 1, 119-127 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Performance Portability and Unified Profiling for Finite Element Methods on Parallel Systems

Vladyslav Kucher[*,1], Jens Hunloh[2], Sergei Gorlatch[2]

[1]*National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Prosp. Peremohy 37, Kyiv, 03056, Ukraine*

[2]*University of Muenster, Einsteinstr. 62, Muenster, Germany*

ARTICLE INFO

Article history:
Received: 11 November, 2019
Accepted: 20 December, 2019
Online: 24 January, 2020

Keywords:
C++ Compilers
GPU programming
High-performance computing
Performance portability
Finite element methods
Parallel computing
Portable profiling
Unified parallel programming

ABSTRACT

The currently available variety of modern, highly-parallel universal processors includes multi-core CPU and many-core GPU (Graphics Processing Units) from different vendors. Systems composed of such processors enable high-performance execution of demanding applications like numerical Finite Element Methods. However, today's application programming for parallel systems lacks performance portability: the same program code cannot achieve stable high performance on different parallel architectures. One of the main reasons for this is that parallel programs are developed by utilizing system-specific profiling interfaces of the corresponding hardware vendors. We describe a novel, portable profiling interface: its design, implementation, and evaluation within the popular framework DUNE for solving differential equations using finite element methods. Our profiler is built on top of the PACXX framework for parallel programming in C++, and it supports portable parallel programming using a single profiling tool on various target hardware platforms.

1 Introduction

Modern universal processors are becoming highly parallel: multi-core CPU and many-core GPU (Graphics Processing Units) are produced by different vendors and exemplify a variety of architecture and hardware solutions. Such processors enable building high-performance systems for computation-intensive applications like numerical finite element methods.

The current programming approaches for parallel computing systems include CUDA [1] that is restricted to GPU produced by NVIDIA, as well as more universal programming models - OpenCL [2], SYCL [3], and PACXX [4] - that follow the idea of unified programming: the programmer can target different hardware without changing the source code, thus reducing the development overhead. However, existing profiling tools are still restricted to the corresponding vendor; therefore, the application programmer usually has to use several different tools to achieve the ultimate portability.

There have been several research efforts to make profiling tools more portable and flexible. CUDA Advisor [5] is a profiling interface that collects data about the performance of both CPU and GPU parts of program code by using the LLVM infrastructure [6]. Experimental approach [7] uses CUDA's TAU tool for profiling GPU applications by producing a detailed information on communications between CPU and GPU, without modifying the source code. The SASSI instrumentation tool (NVIDIA assembly code "SASS Instrumentor") [8] relies on customizable metrics: it enables to inserting user-defined instrumentation code (e.g., debug hooks and customized counters) to collect detailed profiling data that are usually not available. The generic tool integrated in the Score-P infrastructure [9] provides an interface for evaluating the performance of OpenCL code on different parallel architectures.

This paper aims at designing a single, cross-platform profiler that significantly improves the portability of program development. We design our profiler by extending the unified programming framework *PACXX (Programming Accelerators in C++)* [4] with a generic profiling interface: the interface follows the unified programming model of PACXX. Our profiler seamlessly extends the PACXX programming system and enables collecting profiling information at different stages of the development process, for different kinds of the target hardware, thus reducing the development overhead.

[*]Vladyslav Kucher, Prosp. Peremohy 37, Kyiv, 03056, Ukraine, commastergm@gmail.com

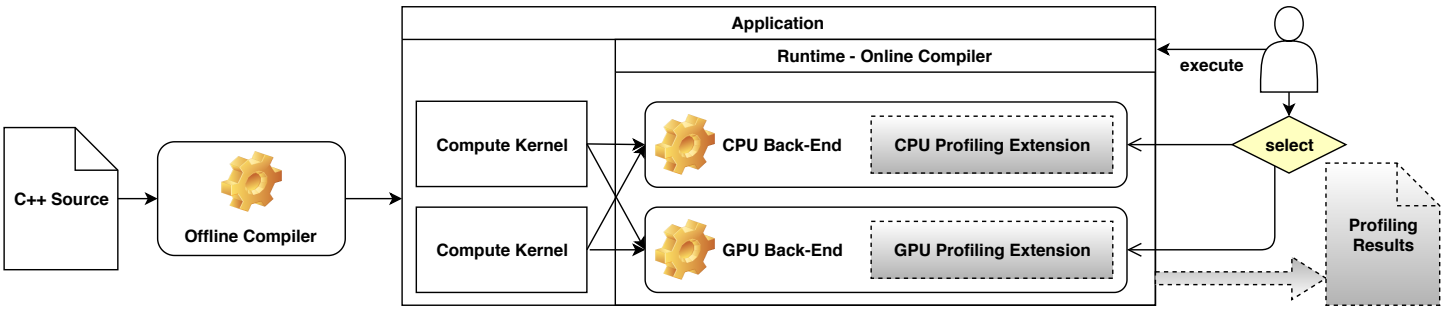


Figure 1: The PACXX framework, extended with profiling (shaded parts)

The unified parallel programming approach of PACXX enables programming CPU/GPU systems in the newest standard C++14/17. PACXX significantly simplifies the development of portable parallel C++ programs by slightly modifying the source code staying within C++. The PACXX advantages were demonstrated for application areas like linear algebra, stencils, and N-Body simulations [10].

We choose *Partial Differential Equations (PDE)* solving as an application field to illustrate and evaluate our approach to unified profiling, because PDE are broadly used in different areas of engineering and science, including simulation [11], finance [12], numerical modeling [13], and biology [14]. Most of practically important PDE are solved using finite-element (FE) methods. Unfortunately, these methods' implementation is often tied to a particular data structure representing the corresponding computation grid. C++ is often used for implementing finite-element methods, because of its high performance and modularity. The *Finite Element Computational Software (FEniCS)* [15] is a C++-based framework for FE variational problems, relying on automatic code generation and the custom *FEniCS Form Compiler* [16]: it supports performance portability by employing a domain-specific language based on Python [17], and multi-staging optimization. The Firedrake approach [18] achieves performance portability by using an advanced code generator: problems are specified in a Python-embedded, domain-specific language that is lowered to C and then JIT-compiled, so target code is generated at run time. The object-oriented framework Overture [19] for solving PDE on various platforms contains C++ classes for computations in different geometries. The Feel++ framework [20] uses a C++-based domain-specific language whose implementation is compatible with third-party libraries.

In order to evaluate our approach to unified profiling, we consider the *Distributed and Unified Numerics Environment (DUNE)* [21] as a case study. DUNE offers a unified interface for solving PDE on parallel systems [22]: the interface combines various methods in a modular C++ library. DUNE enables using arbitrarily shaped grids and a large set of particular algorithms. While the design of DUNE is very flexible, its application performance strongly depends on the employed C++ compiler and the target hardware.

The goal of this paper is to extend the flexibility of existing approaches to parallel PDE solving. We achieve performance portability and cross-platform profiling of the finite-element methods for solving PDE, by extending the PACXX framework and integrating it with the existing DUNE framework.

The paper is organized as follows. In Section 2, we describe the C++-based unified programming approach of PACXX and the design of our profiling interface. In Section 3, we show how this interface is used for a portable profiling a simple case on various target architectures. In Sections 4 and 5, we illustrate how PACXX can accelerate grid-based FE algorithms for PDE in the extended DUNE. We demonstrate that the integration of PACXX with DUNE leads to performance portability over different architectures, significantly decreasing the development overhead and improving code maintenance. We experimentally evaluate our unified profiling approach in Section 6, and we summarize our results in Section 7.

2 Programming and profiling

2.1 The PACXX Programming model

We use the PACXX framework [4] that supports a unified, C++-based model of parallel programming. In analogy with the recent OpenCL standard [23], PACXX uses the parallelism expressed using kernels that are executed on devices. The main advantage of PACXX, however, is that an application is expressed as single-source C++ code, whereas OpenCL kernels are implemented in a special kernel language and they need additional host code to be executable. The performance portability in PACXX is ensured by means of several pre-implemented back-ends.

Figure 1 shows that the compilation process of PACXX proceeds in two steps. The first step is the offline compilation that transforms source code into an executable. Kernel's code is precompiled, and the executable is integrated with the particular back-end for the target CPU or GPU. In the second stage, online compilation is invoked when the executable is started. The integration of the executable with back-ends for different architectures allows the user to choose the target hardware for each execution. The efficient execution on the chosen target system is handled by PACXX transparently: all necessary generation steps and optimizations are performed automatically. This brings the portability of PACXX applications on CPUs and GPUs of different vendors.

2.2 Profiling interface: Design and implementation

Figure 1 shows the design of our unified profiling interface integrated into PACXX: shaded are our profiling extensions to the original PACXX back-ends. These extensions comprise hardware-specific code to collect profiling data. For the user, our profiling interface can be viewed as a wrapper that can choose the corresponding profiling extension for the particular target architecture.

The advantage of our design of profiling is that no changes to the offline compiler of PACXX are necessary. The modifications are made only to the PACXX runtime and back-ends: they are extended with vendor-specific profiling tools: the CPU back-end uses the PAPI library [24], and the back-end for NVIDIA GPUs relies on the CUPTI library [25].

2.3 The profiling extension

For different back-ends, the profiling extensions have a unified structure that includes a changed launch procedure for kernels to manage collecting performance data. We have to modify the way how a kernel is launched, because many existing devices cannot record several profiling metrics simultaneously (e.g., NVIDIA GPUs [25]). For some metrics, multiple reproducible runs of the same kernel may be necessary to record that metric. It becomes mandatory to rebuild the device memory state automatically after each kernel execution, such that the executions are measured independently. This reconstruction of the memory state enables multiple profiling runs for the same program, so that accurate performance data are produced. Traditionally, this is done using a shadow copy mechanism that stores and restores the device memory state transparently at each kernel execution. This solution effectively decouples the profiling from undesirable side effects: eventually this ensures exact performance data and deterministic behavior when measuring multiple kernel launches.

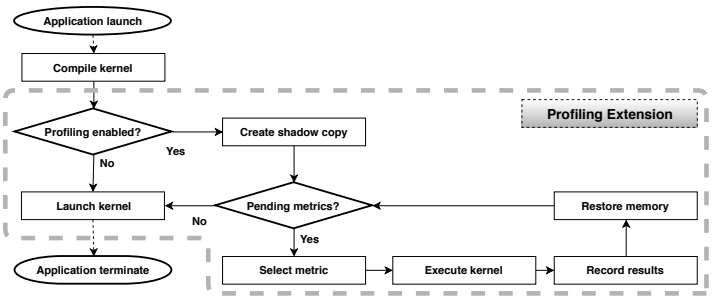


Figure 2: Kernel launch, modified for profiling

Figure 2 shows our modification (in dashed box) to the launch procedure within a PACXX back-end. The new launch procedure reflects the control flow of our profiling extension: the processing steps rely on target-specific code of the concrete back-end. When profiling is enabled, a kernel launch triggers the profiling extension to store a shadow copy of the utilized part of device memory. Then the kernel is executed repeatedly, such that this shadow copy is used for every recorded performance metric. Eventually, the application proceeds using the original kernel launch procedure.

3 The profiler usage

We demonstrate the usage of and evaluate our unified profiler on two application examples - the first is traditional multiplication of matrices, and the second is a more elaborated PDE solver.

3.1 Example: Matrix multiplication

Figure 3 illustrates how matrix multiplication is expressed in PACXX [4]: this program is created from the C++ code by transforming nested sequential loops into parallel calls of a kernel. Arrays *a* and *b* stand for the input matrices, and array *c* stores the resulting matrix. The *matrixMultKernel* kernel (lines 15-23) is written as a C++ lambda expression to be computed in parallel for each element of *c*. The kernel run requires an upload of the input data (lines 3-13) and the result fetching (line 26) after the computation, which is analogous to CUDA and OpenCL kernels.

```
1  auto& device = Executor::get(0);
2
3  auto& da = device.allocate<double>(matrix_size);
4  auto& db = device.allocate<double>(matrix_size);
5  auto& dc = device.allocate<double>(matrix_size);
6
7  da.upload(a, matrix_size);
8  db.upload(b, matrix_size);
9  dc.upload(c, matrix_size);
10
11 auto pa = da.get();
12 auto pb = db.get();
13 auto pc = dc.get();
14
15 auto matrixMultKernel = [=](auto &config)
16 {
17    auto col = config.get_global(0);
18    auto row    = config.get_global(1);
19    double val  = 0;
20    for (unsigned i = 0; i < width; ++i)
21       val += pa[row * width + i] *
                pb[i * width + col];
22    pc[row * width + col] = val;
23 };
24 exec.launch(matrixMultKernel,
     {{width/threads , width} , {threads , 1} , 0});
25
26 dc.download(c, matrix_size);
```

Figure 3: PACXX code example: matrix multiplication

Figure 3 shows that the code employs parallelism by dividing calculations in blocks and threads. Function *launch* (line 24) invokes the kernel and it partitions work in dimensions (up to three), thus specifying the structure and degree of parallelism. In the range, the first component declares how many blocks are started, while the *threads* variable in the range declares the threads number in a block. At run time, each block is assigned to a processor that runs all threads of the block in the SIMT (Single Instruction, Multiple Threads) manner, which is a combination of SIMD and multi-threading paradigms. Function *get_global* retrieves the thread id in each dimension of the range. The first dimension of the range in our example corresponds to the row, and the second to the column in the matrix.

After the invocation of PACXX offline compiler, the executable is automatically combined with our profiler. Therefore, PACXX serves effectively as a drop-in for the standard LLVM (Low-Level Virtual Machine) [6] compilation toolchain.

3.2 Workflow of profiling

In Figure 4, the profiling of our example application is invoked with the runtime environment variable PACXX_PROF_ENABLE enabled, which leads to profiling each kernel invocation. At each kernel run, the runtime system executes the online compilation transparently for the user and it enables the profiler, such that the profiling information is collected from the target device. Variable PACXX_DEFAULT_RT specifies the target architecture (CPU or GPU). Profiling data are written in an output file that can be either explicitly specified as in Figure 4 or sent to the standard output.

```
> PACXX_DEFAULT_RT=GPU
  PACXX_PROF_ENABLE=1
  PACXX_PROF_IN=custom_profiling_configuration
  PACXX_PROF_OUT=pacxx.json
  ./matrixMult
```

Figure 4: Invoking a PACXX program with profiling

Our profiler always reports one metric - runtime duration - by default. To ask for additional profiling data, the user can specify a specific configuration. The profiler is able to profile all metrics that are supported by the system vendor, e.g., memory throughput and usage, instructions per cycle, etc. All 174 possible metrics for NVIDIA GPUs are listed in [25].

3.3 Visualization of profiling

Figure 5 illustrates the profiling data collected for the matrix multiply example prepared as shown in Figure 4. The profiler produces a JSON-format file that comprises all specified metrics for each started kernel instance. For multiple kernel launches, our profiler records every launch independently, and the output data are prefixed by the kernel name. Thus, the user obtains profiling information automatically: the user does not have to explicitly select the profiling tool for a kernel and the architecture on which it runs: the appropriate profiling functionality is ensured by the PACXX runtime for the target architecture.

```
"matrixMultKernel": [
  { "Metrics": {
      "cf_executed": "68157540",
      "kernelDuration": "517349264ns" } } ]
```

Figure 5: Results of profiling in JSON format

The use of the JSON format enables efficient visualization of the profiling results using Gnuplot or other tools, such that the code behavior on different target platforms can be conveniently compared.

Figure 6 compares the execution time of matrix multiplication on a CPU (Intel Xeon E5-1620 v2) vs. a GPU (NVIDIA Tesla K20c). The input comprises two 4096x4096 square matrices generated randomly. The plot demonstrates how the execution time is dependent of the parallelism amount specified by parameter *threads* in the code in Figure 3.

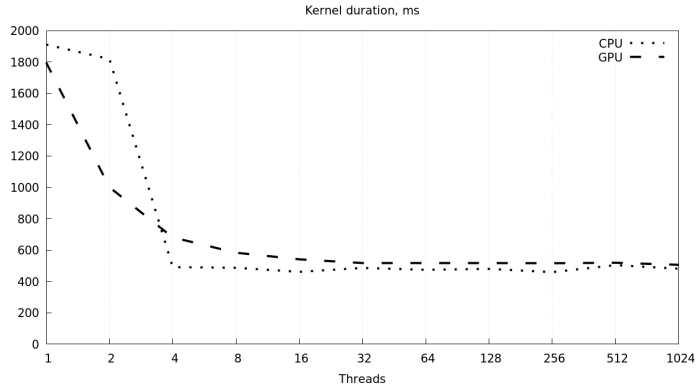


Figure 6: Execution time of matrix multiply on CPU vs. GPU

Figure 7 compares the number of branch statements on CPU vs. GPU, depending on the number of threads.

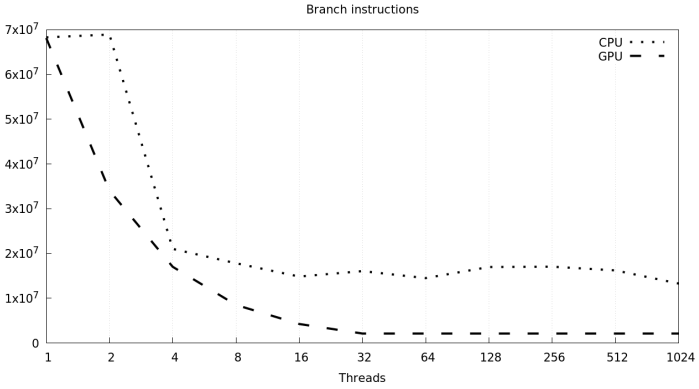


Figure 7: Number of branch instructions on CPU vs. GPU

Summarizing, our profiler enables both specifying metrics and representing profiling results in a uniform manner, such that different target architectures can be compared with each other conveniently.

4 Integrating DUNE with PACXX

In this section, we show the use of our approach to profiling on the case study of the DUNE framework. We explain its efficient integration with the PACXX unified programming framework, by taking into account the design of DUNE's grid interface.

Our idea of integration is to employ PACXX as a code generator for the DUNE interface which is designed by using abstract interfaces and template meta-programming to separate data structures and algorithms. By reusing existing modules, DUNE offers a rich collection of functionalities.

In our integration with PACXX, we use the fact that DUNE contains C++ modules that provide abstract interfaces with several implementations. The use of generic programming [26] and C++ template meta-programming [27] enables optimizations to be applied at compile time: therefore, DUNE is flexible, without introducing overhead at runtime.

Figure 8 illustrates that our integration comprises two levels. At the top (abstraction) level, a particular problem is formulated

using extensions of a domain-specific language analogous to [28], and is then implemented as a specific DUNE kernel. At the bottom (hardware optimization) level, code is generated for a specific target architecture. This two-level design of the integration toolchain enables suitable optimizations at every processing level.

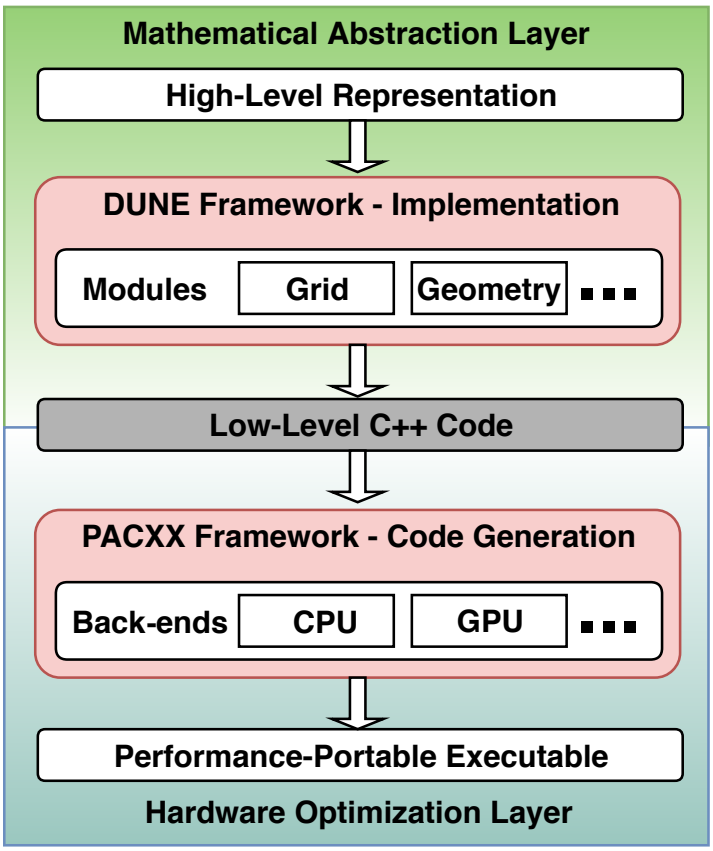


Figure 8: Integration of PACXX with DUNE

The integrated toolchain is now as follows. The PDE problem is stated by the domain expert as an expression of a domain-specific language, which is then implemented by a specific grid-based kernel of DUNE: the framework yields C++ code that contains the resulting kernel with template arguments and runtime values. This code is eventually compiled by PACXX, as we explain below.

The PACXX framework employs the popular LLVM compiler infrastructure and its Intermediate Representation (IR) [6] to obtain portable executables with hardware-specific optimizations. Kernels are compiled offline to IR and then the online compiler generates code and optimizes it for the target architecture [10] transparently for the user. This toolchain enables portable PACXX applications over architectures of modern parallel processors. PACXX also supports C++ template meta-programming which significantly reduces the runtime overhead, and our unified profiling allows profiling of the DUNE without further modifications to the code.

The main advantage of using PACXX vs. existing profilers is the profiling of applications over various platforms, without having to use different tools. Our profiler improves the flexibility of programming process: the program can be configured regarding requested profiling data on every supported target architecture. In addition, due to the use of PACXX for generating code in DUNE we can apply optimizations based on control-flow linearization [29]. Since PACXX is fully compatible with the modern C++14/17 standards, the program developers can use existing C++ code on various target architectures, which significantly reduces the development effort.

5 Compiling kernels: examples

This section describes the use of our integrated DUNE-PACXX framework for generating code for two examples of DUNE kernels: 1) Poisson equation solving, and 2) deformation of material calculation based on the linear elasticity theory. These two examples cover two scenarios: Poisson calculates with low load on a simple 2D grid, and linear elasticity is based on a 3D grid with a heavy load. Example applications of linear elasticity are in computer-aided design, and Poisson is applied in physics to calculate fields of potentials.

5.1 PACXX parallelization

We consider example kernels that are simple sequential C++ codes; they are parallelized according to the PACXX programming approach. The slight one-time adaptation effort of the kernels leads to the performance portability of the resulting code. Thereby the programmer does not have to re-design the program code for each new parallel architecture.

Our two kernel examples work on a grid refined on two levels: the computation in a uniform grid of *macro-elements* uses a coarse-grained refinement, and each macro-element is further refined to a grid of *micro-elements*. We parallelize kernels by two ways: either different macro-elements are processed in parallel, or several micro-elements in a macro-element are processed in parallel.

We compare the micro- vs. macro-parallelization approaches regarding their performance against an original sequential version.

5.2 The Poisson kernel

```
1  // iterating over micro-elements
2  for (int el_y=0; el_y<k; ++el_y)
3   for (int el_x=0; el_x<k; ++el_x)
4   {
5    // (1) calculating
6    RF grad_u[2] = {};
7    for (int dim=0; dim<=1; ++dim)
8    {
9     for (int index_y=0; index_y<2; ++index_y)
10     for (int index_x=0; index_x<2; ++index_x)
11      grad_u[dim] += grad[index_x+index_y*2][0][dim]
        * x(lfsv, (el_y+index_y)*(k+1) + el_x + index_x);
12   }
13   // (2) accumulating
14   for (int index_y=0; index_y<2; ++index_y)
15    for (int index_x=0; index_x<2; ++index_x)
16     r.accum(lfsv,
       (el_y+index_y)*(k+1) + el_x + index_x,
       factor*(grad[index_x+index_y*2][0][0]*grad_u[0]
              +grad[index_x+index_y*2][0][1]*grad_u[1]));
17 }
```

Figure 9: Poisson kernel: sequential C++ code

Figure 9 illustrates the C++ code for the Poisson kernel that computes a single macro-element in DUNE. Two nested *for* loops in

lines 2-3 iterate on the micro-grid. For every element, the computation comprising the loop body is executed. Local basis functions $lfsv$ and x in line 11 compute a gradient, and the consequent accumulation phase (line 16) uses the calculated gradient for updating the computational grid.

Figure 10 shows how the memory is accessed by the kernel in Figure 5.

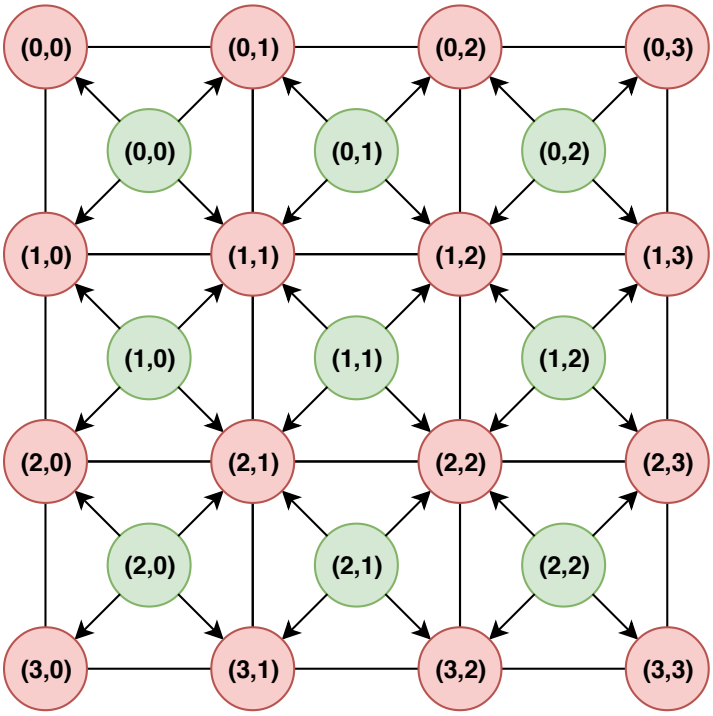


Figure 10: Poisson: micro-grid and data dependencies

For correct parallelization, we must avoid race conditions. Therefore, we synchronize accesses to memory by distinguishing between two kernels - accumulation and computation. We parallelize the calculation kernel on the micro-grid elements, but temporarily store the results of the calculation kernel, rather than performing write operations immediately.

Afterwards, the corresponding vertices are processed in parallel by the accumulation kernel: it reads the previously stored intermediate results and updates the adjacent vertices by simultaneous write operations. This described implementation guarantees the absence of conflicts in the patterns of memory access at the level of micro-grid.

At a macro-layer, we parallelize computations by executing the main computation for each instance of the kernel, such that process in parallel all macro-elements. For this approach, DUNE's grid interface must be slightly modified, but the advantage is that the patterns of memory access are not so complicated.

The parallelization at macro-layer implies that workload is distributed in a coarse-grained manner; this is advantageous on large-scale since it greatly reduces the parallelization overhead. However, it brings a drawback: global memory must be allocated by all macro-elements simultaneously, so the memory requirement increases. Vice versa, the fine-grained distribution of workload for micro-layer parallelism allocates little memory for a macro-element at a time. Therefore, large micro-grids with enough compute load in a single macro-element especially benefit from the micro-layer parallelization.

5.3 Kernel for linear elasticity

Figure 11 shows our second example - the linear elasticity kernel – that consists of three nested *for* loops. and is thus significantly more complex than Poisson. On the one hand, this complicates the parallelization (while the scheme of parallelization remains similar to Poisson), but on the other hand, this offers more potential of parallelism.

```
1  // iterating over micro-elements
2  for (int el_z = 0; el_z < k; ++el_z)
3   for (int el_y = 0; el_y < k; ++el_y)
4    for (int el_x = 0; el_x < k; ++el_x)
5    {// (1) calculating
6     for (int dim = 0; dim <= 2; ++dim)
7      for (int index_z = 0; index_z <= 1; ++index_z)
8       for (int index_y = 0; index_y <= 1; ++index_y)
9        for (int index_x = 0; index_x <= 1; ++index_x)
10       {// computating the element values ...}
11    // (2) accumulating
12    for (int index_z = 0; index_z <= 1; ++index_z)
13     for (int index_y = 0; index_y <= 1; ++index_y)
14      for (int index_x = 0; index_x <= 1; ++index_x)
15       {// accumulation of node values...}
```

Figure 11: Linear elasticity: sequential C++ code

6 Experimental evaluation

We conduct our measurements using two parallel processors: an Intel CPU Xeon E5-1620v2 with 4 cores of frequency 3.70GHz (8 logical cores), and an NVIDIA GPU Tesla K20c with 2496 compute units. Our CPU contains AVX with vector registers, so we evaluate both scalar and auto-vectorized versions of our kernels. The number of (sub)elements (x-axis) corresponds to the subdivision degree of the grid in each dimension: e.g., 5 sub-elements imply a 5x5x5 grid in a 3D program for linear elasticity.

Each measurement performs one iteration of the linear elasticity program which corresponds to 16 executions of the kernel based on the amount of quadrature points. The following figures compare the three aforementioned platforms (vectorized CPU, scalar CPU and GPU) based on various metrics.

Figure 12 and Figure 13 show the memory operations (lower is better). We observe that read operations can be coalesced quite well, while write operations less so. The GPU with its 32 memory controllers sets the target for other hardware platforms which they struggle to reach.

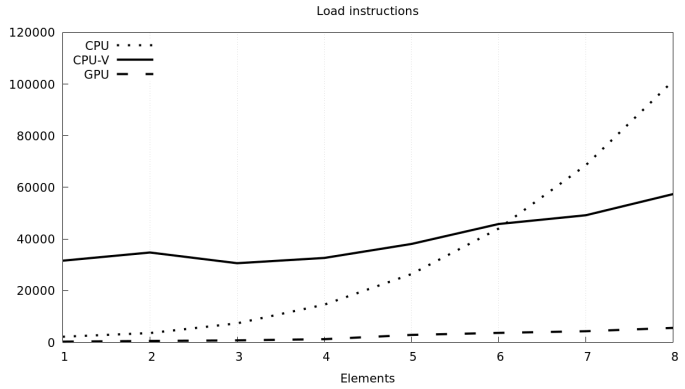


Figure 12: Number of executed memory reads across used platforms

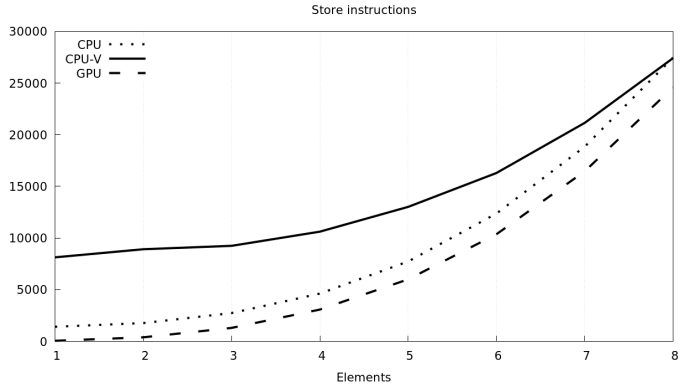


Figure 13: Number of executed memory writes across used platforms

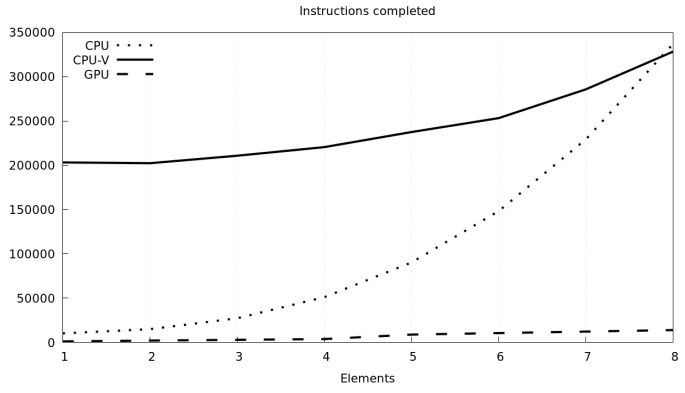


Figure 14: Total completed instructions across used platforms

On the code side, Figure 14 shows the number of executed instructions. We see that CPU vectorization suffers from the low data sizes generating too many scatter-gather instructions (Figure 15 further supports this claim, showing an order of magnitude more processed control-flow instructions in the vectorized CPU version of the kernel). The GPU can distribute the load across its numerous 2496 computing units and keep the number of instructions low.

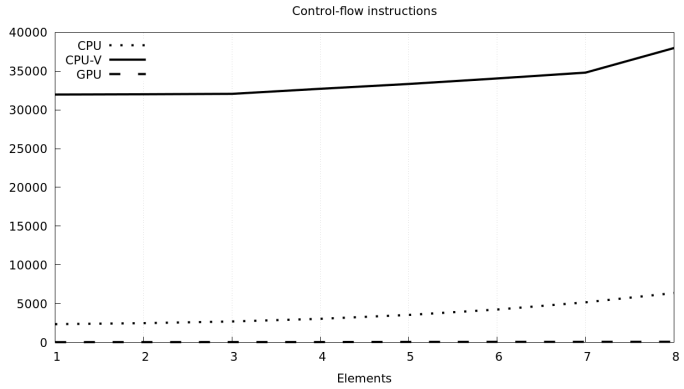


Figure 15: Number of executed control-flow instructions across used platforms

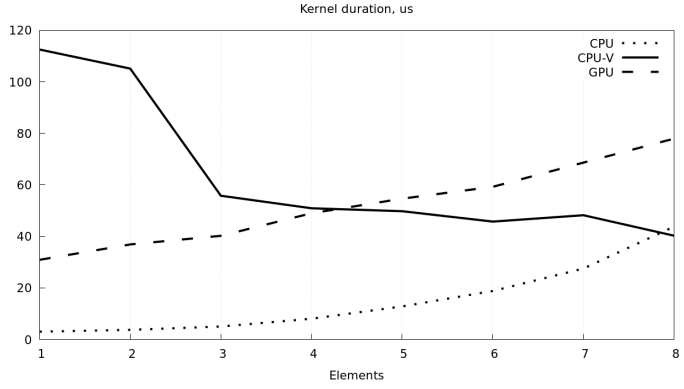


Figure 16: Linear elasticity kernel duration across used platforms

Figure 16 compares the run time of kernel execution. Vectorized CPU version of the kernel does not show good performance with low data amounts, but it improves over the scalar CPU version by element count of 8, matching with the expectation from Figure 14. Surprisingly, the GPU version lags behind the scalar CPU version, so we look into the reasons of low performance on the GPU side.

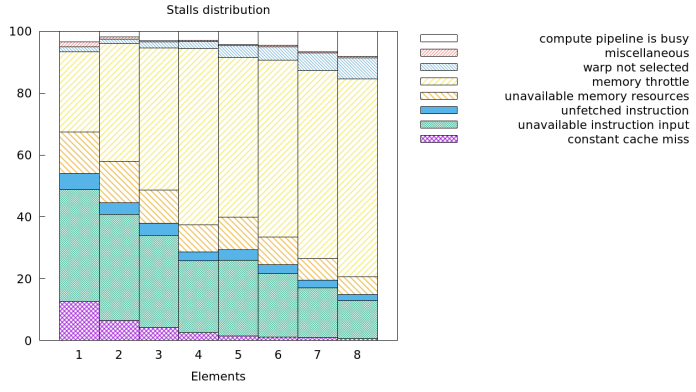


Figure 17: Execution stalls on the GPU causing low performance

Figure 17 shows that the major slowdown reason is memory throttling caused by the hardware limitation to issue memory instructions every 4 cycles. Therefore, memory requests have high divergence and, thus, cannot be fulfilled in a timely manner.

7 Conclusion

We design and implement a novel cross-platform approach to programming and profiling parallel applications, which significantly improves the portability of program development. We extend the unified programming framework PACXX with a generic profiling interface that enables collecting profiling information at different stages of the development process, for different target hardware. Our approach liberates the program developer from having to use several proprietary tools when profiling the same application on different hardware. Therefore, the software development overhead is significantly reduced.

We choose Partial Differential Equations (PDE) solving to illustrate and evaluate our approach of unified profiling, because PDE are broadly used in different areas of engineering and science. As a case study, we use the popular DUNE framework, and we experimentally confirm that the integration of our extended PACXX with DUNE leads to performance portability over different architectures, significantly decreasing development overhead and improving code maintenance.

Acknowledgement The authors gratefully acknowledge generous support from the German Federal Ministry of Education and Research (BMBF) within the HPC^2SE project.

References

[1] J. Nickolls, I. Buck, M. Garland, K. Skadron, "Scalable Parallel Programming with CUDA," *Queue*, vol. 6, pp. 40–53, Mar. 2008.

[2] J. Kim, S. Seo, J. Lee, J. Nah, G. Jo, J. Lee, "OpenCL As a Unified Programming Model for Heterogeneous CPU/GPU Clusters," *SIGPLAN Not.*, vol. 47, pp. 299–300, Feb. 2012.

[3] R. Keryell, R. Reyes, L. Howes, "Khronos SYCL for OpenCL: A Tutorial," in *Proceedings of the 3rd International Workshop on OpenCL*, IWOCL '15, (New York, NY, USA), pp. 24:1–24:1, ACM, 2015.

[4] M. Haidl, M. Steuwer, T. Humernbrum, and S. Gorlatch, "Multi-stage Programming for GPUs in C++ Using PACXX," in *Proceedings of the 9th Annual Workshop on General Purpose Processing Using Graphics Processing Unit*, GPGPU '16, (New York, NY, USA), pp. 32–41, ACM, 2016.

[5] D. Shen, S. L. Song, A. Li, X. Liu, "CUDAAdvisor: LLVM-based Runtime Profiling for Modern GPUs," in *Proceedings of the 2018 International Symposium on Code Generation and Optimization*, CGO 2018, (New York, NY, USA), pp. 214–227, ACM, 2018.

[6] C. Lattner, V. Adve, "LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation," in *Proceedings of the International Symposium on Code Generation and Optimization: Feedback-directed and Runtime Optimization*, CGO '04, (Washington, DC, USA), pp. 75–, IEEE Computer Society, 2004.

[7] A. D. Malony, S. Biersdorff, W. Spear, S. Mayanglambam, "An Experimental Approach to Performance Measurement of Heterogeneous Parallel Applications Using CUDA," in *Proceedings of the 24th ACM International Conference on Supercomputing*, ICS '10, (New York, NY, USA), pp. 127–136, ACM, 2010.

[8] M. Stephenson, S. Hari, S. Kumar, Y. Lee, E. Ebrahimi, D. R. Johnson, D. Nellans, M. O'Connor, and S. W. Keckler, "Flexible Software Profiling of GPU Architectures," *SIGARCH Comput. Archit. News*, vol. 43, pp. 185–197, June 2015.

[9] R. Dietrich, R. Tschüter, "A generic infrastructure for OpenCL performance analysis," in *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 1, pp. 334–341, Sept 2015.

[10] M. Haidl, B. Hagedorn, S. Gorlatch, "Programming GPUs with C++14 and Just-In-Time Compilation," in *Parallel Computing: On the Road to Exascale, Proceedings of the International Conference on Parallel Computing, ParCo 2015, 1-4 September 2015, Edinburgh, Scotland, UK*, pp. 247–256, 2015.

[11] F. Feldmann, B. Hagemann, L. Ganzer, M. Panfilov, "Numerical simulation of hydrodynamic and gas mixing processes in underground hydrogen storages," *Environmental Earth Sciences*, vol. 75, p. 1165, Aug 2016.

[12] Daniel J. Duffy, *Finite Difference methods in financial engineering: a Partial Differential Equation approach*. John Wiley & Sons, 2013.

[13] I. Rucker, W. Ressel, "A numerical drainage model to simulate infiltration into porous pavements for higher road safety," in *17. Internationales Stuttgarter Symposium*, (Wiesbaden), pp. 1293–1303, Springer Fachmedien Wiesbaden, 2017.

[14] Anthony W. Leung, *Systems of nonlinear partial differential equations: applications to biology and engineering*, vol. 49. Springer Science & Business Media, 2013.

[15] M. S. Alnæs, J. Blechta, J. Hake, *et al.*, "The FEniCS project version 1.5," *Archive of Numerical Software*, vol. 3, no. 100, pp. 9–23, 2015.

[16] A. Logg, K. B. Ølgaard, M. E. Rognes, and G. N. Wells, "FFC: the FEniCS form compiler," in *Automated Solution of Differential Equations by the Finite Element Method*, pp. 227–238, Springer, 2012.

[17] G. R. Markall, F. Rathgeber, L. Mitchell, *et al.*, "Performance-Portable Finite Element Assembly Using PyOP2 and FEniCS," in *Supercomputing*, (Berlin, Heidelberg), pp. 279–289, Springer Berlin Heidelberg, 2013.

[18] F. Rathgeber, D. A. Ham, L. Mitchell, *et al.*, "Firedrake: automating the finite element method by composing abstractions," *ACM Transactions on Mathematical Software (TOMS)*, vol. 43, no. 3, p. 24, 2017.

[19] D. L. Brown, W. D. Henshaw, and D. J. Quinlan, "Overture: An object-oriented framework for solving partial differential equations," in *International Conference on Computing in Object-Oriented Parallel Environments*, pp. 177–184, Springer, 1997.

[20] C. Prud'Homme, V. Chabannes, V. Doyeux, *et al.*, "Advances in Feel++: a domain specific embedded language in C++ for partial differential equations," in *Eccomas' 12-European Congress on Computational Methods in Applied Sciences and Engineering*, 2012.

[21] P. Bastian, M. Blatt, C. Engwer, *et al.*, "The distributed and unified numerics environment (DUNE)," in *Proc. of the 19th Symposium on Simulation Technique in Hannover*, vol. 123, 2006.

[22] P. Bastian, M. Blatt, A. Dedner, *et al.*, "A generic grid interface for parallel and adaptive scientific computing. Part I: abstract framework," *Computing*, vol. 82, no. 2-3, pp. 103–119, 2008.

[23] J. E. Stone, D. Gohara and G. Shi, "OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems," *IEEE Des. Test*, vol. 12, pp. 66–73, May 2010.

[24] J. Dongarra, K. London, S. Moore, P. Mucci and D. Terpstra, "Using PAPI for Hardware Performance Monitoring on Linux Systems," 08 2009.

[25] "CUPTI API." https://docs.nvidia.com/cuda/cupti/r_main.html. Accessed: 2018-07-14.

[26] P. Bastian, F. Heimann, and S. Marnach, "Generic implementation of finite element methods in the distributed and unified numerics environment (DUNE)," *Kybernetika*, vol. 46, no. 2, pp. 294–315, 2010.

[27] M. Blatt and P. Bastian, "The iterative solver template library," in *International Workshop on Applied Parallel Computing*, pp. 666–675, Springer, 2006.

[28] M. S. Alnæs, A. Logg, K. B. Ølgaard, M. E. Rognes, and G. Wells, "Unified Form Language: A Domain-specific Language for Weak Formulations of Partial Differential Equations," *ACM Trans. Math. Softw.*, vol. 40, pp. 9:1–9:37, Mar. 2014.

[29] M. Haidl, S. Moll, L. Klein, H. Sun, S. Hack, and S. Gorlatch, "PACXXv2+RV: An LLVM-based Portable High-Performance Programming Model," in *Proceedings of the Fourth Workshop on the LLVM Compiler Infrastructure in HPC*, LLVM-HPC, (New York, NY, USA), pp. 7:1–7:12, ACM, 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 128-134 (2020)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Sensor Based on-the-go Detection of Macro Nutrients for Agricultural Crops

Shraddha Balayya Hiremath*, Raghavendra Shet, Nikita Patil, Nalini Iyer

School of ECE, KLE Technological University, 580031, India

ARTICLE INFO

Article history:
Received: 31 May, 2019
Accepted: 02 November, 2019
Online: 22 January, 2020

Keywords:
Agriculture
Macro nutrients
SPAD meter

ABSTRACT

Agriculture, being the vital sector, contributes to India's GDP by 18%. It is necessary to enhance the production in this sector with minimal need of resources to get the good yield of crops. It in turn boosts up the necessity of automation in the field of agriculture. This paper discusses about the experimental study which includes the invasive/non-invasive method to sense the approximate supply of macro nutrients such as Nitrogen, Phosphorous & Potassium to the plants. The nutrient content of leaf is sensed by the usage of optical sensors. The results have been obtained based on the sensor data carried out for the sample test leaf. Thus, the level of the estimated nutrient contents is indicated by three different ranges such as Low (L), Medium (M) and High (H). A mathematical model is developed for nutrient measurement especially for nitrogen and validated with SPAD meters for the measurement of nitrogen content. The other macro nutrients were also estimated, and the entire system is tested at different test conditions with the support of UASD in the field of University of Agricultural Science.

1. Introduction

As there is an increase in population and the per capital income, the demand for crop yield is drastically increasing with the advancement in technological growth. It is necessary to identify the severe problems involved in crop yield and address the challenges [1]. Even though there is much advancement in the field of science and technology still farmers across the country adopt the traditional practices. The availability of the resources in the process of farming is one of the most important aspects in eco-system. In addition to this nutrient are very important resources for the plants that give structure to the leaf. Since the population growth of the nation is getting enhanced, this creates a necessity to increase the food production. Supplying the appropriate amount of fertilizers and water becomes evident in enhancing the plantation yield as a whole. The basic nutrients provided to the plants are Nitrogen (N), Phosphate (P), and Potassium (K) in the ratio of 4:2:1. On the other hand due to the lack of awareness and quick measurement techniques, the farmers oversupply nutrients and water to plants, leading to destruction of the crops. Especially the excessive supply of nitrogen nutrient to the crops causes the chemical to evaporate and create oxides of nitrogen (NOx) mainly nitric oxide and nitrogen dioxide into the atmosphere. The formed NOX contributes to the formation of smog and acid rain, affecting ozone layer resulting to global warming. Hence there is a necessity for efficient use of nitrogen broadly known as Nitrogen Use Efficiency (NUE).

The fertilizers should be supplied in appropriate manner so that good number of crops can be obtained without any defects. The exact amount of fertilizers will be supplied based on the type of crop, its growth status and the quality of fertilizers or soil. The growth of the plant depends upon absorption and consumption of the nutrients. One of the major challenges in agriculture in order to get the optimized plant yield is reducing the amount of water consumption and fertilizers [2]. Different nutrients are available in the market for plantations. They are scattered to the plantation area for better crop output. Nitrogen related chemicals are distributed in abundant volume as they play an active role in managing an edulcorated premises in the field. Also, potassium and phosphorous aids in plant growth. The growth stage of the plant can be managed in a better way by maintaining proper supply of these nutrients. But many times, unnecessary supply or inadequate supplies of nutrients are evident in all field area leading to environmental hazards and wastage of crop yield. In association with the technology, a smart way has to be established to accurately measure the nutrient content and supply concrete chunk of fertilizers.

*Shraddha Balayya Hiremath, Email: shraddha_h@bvb.edu

The conventional methods of Chlorophyll and Nitrogen analysis require the destruction of the leaf sample. It is also time-consuming and expensive process. The above techniques cannot satisfy the requirements of precision farming. Non-destructive optical techniques based on absorbance and reflectance of light by leaves have been proven as alternative time-saving and simple techniques to quantify Chl in a number of agricultural species such as birch (*Betula pendula Roth.*)[3], [4], wheat (*Triticum aestivum L.*)[5] and potato (Solanum tuberosum L.), maize (*Zea mays L.*)[6] and soybeans (*Glycine max L. Merrill*.), sorghum (*Sorghum bicolor L. Moench*.) and pigeonpea (*Cajanus cajan L. Millsp*.). Strong relationship between CCI and nitrogen was also reported in sugar maple (*Acer saccharum Morsh*.) (Van den Berg and Perkins2004) [7], [8], and rice (*Oryza sativa L*).

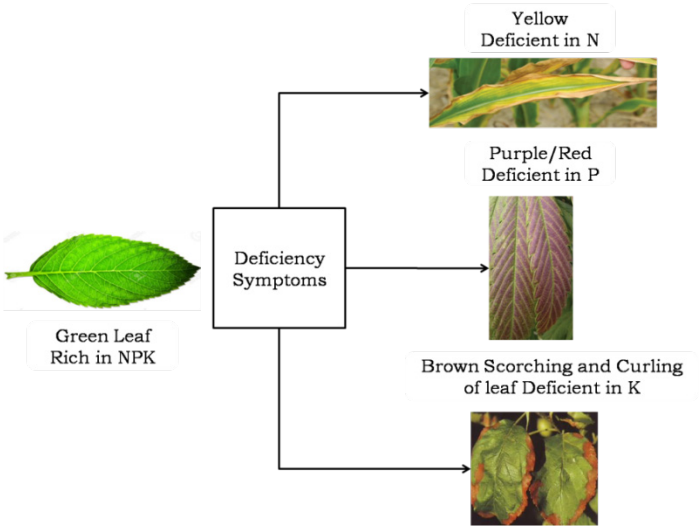


Figure 1: Change of color showcasing nutrient deficiency in leaf

From Figure 1, it is apparent that the deficiency in nutrients results in an outcome in the form visual look of the leaf which proves to be a nutrient deficient plant [9]. The deficiency symptoms in plants completely depend upon the color of the leaf. They are as follows,

- If the color of the leaf results in yellow, it proves to be deficient in N
- If the color of the leaf results in purple/red, it proves to be deficient in P
- If the leaf results in brown scorching and curling of leaf, it proves to be deficient in K

Finally, if the leaf appears to be completely green at its initial and later stages of growth, it results to be rich in all the three macro nutrients.

Methodology carried out to detect the approximate value of nutrients is by the study of spectral reactance of the leaf of the plant. The reactance alters at different stages of the plant growth on the grounds of variation in nutrient content [3]. In particularly the nitrogen is absorbed by the plant through atmosphere and is closely proportional in a linear fashion with chlorophyll. Hence the penetration and reactance of the leaf with light modulates with variation of the color of the leaf from green to yellow. On the other way the different macro nutrients are directly sourced from the soil of the plants.

The techniques present for nutrient and moisture content in plant are:

- Invasive Method
- Non-invasive method

Invasive technique uses destructive method for measurement of nutrients. The invasive methods used for measuring micro and macro nutrients for soil, leaf and irrigation water are:

- Electro-chemical method.
- Ion-selective electrode method.

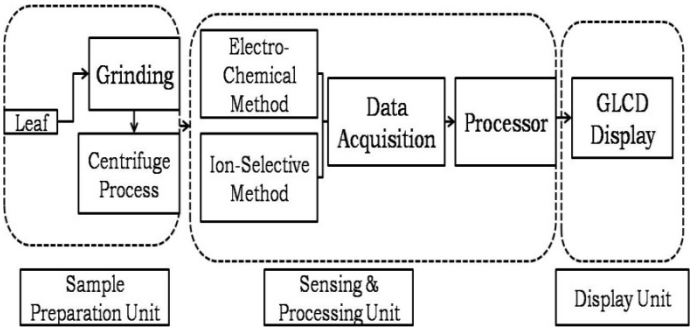


Figure 2: Architecture for Invasive Method

The overall system level block diagram for Invasive technique is depicted in Figure 2. It comprises of the various blocks such as sample preparation unit, sensing & processing unit and display unit. The sample preparation unit consists of sample leaf which is grinded into small pieces and further sent to the centrifuge process where in the sedimentation principle is used to separate the denser substances from less dense ones. The sensing and processing unit inculcates two different methods such as electro-chemical method and ion-selective method to acquire the data from previous unit. The data is further processed and given to the display unit in order to display the required data. The architecture for Non-invasive method is shown in Figure 3. The non-invasive technique uses non-destructive method for measurement of nutrients. This method uses non-contact type of sensors. The optical based sensors with the appropriate wavelength are selected for measuring the nutrients. The output obtained from optical sensor is processed using an embedded system and displayed. Based on the concentration of the nutrients an automatic alarm signal is generated for an inadequate concentration of particular nutrients which in turn controls the switching ON and OFF of the existing drip irrigation system.

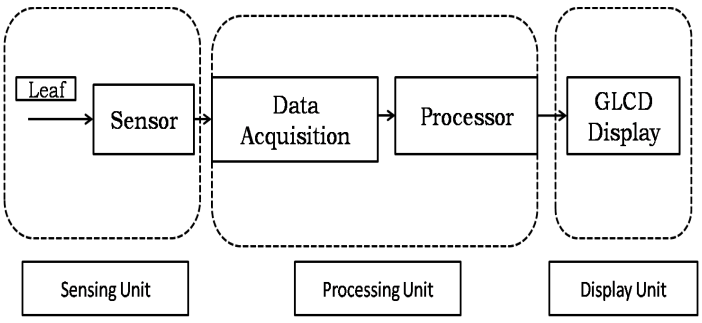


Figure 3: Architecture for Non-invasive Method

In order to quantify the spectral absorption optical sensor was used termed as TCS230.A waveform is generated as a result of the sensor which is later sampled to quantify the changes in nitrogen nutrient in leaf [10]. Different plant families exhibit variations mathematically involving Chlorophyll content Index (CCI) and entire quantity of chlorophyll/nitrogen content [11], [12]. For the product under design a self reliant mathematical model is developed for the species under consideration to a belonged zone. The main feature of the system should include its portability and user friendly behavior.

The organization of the paper is as follows. Section II discusses the implementation details of the system involving nitrogen, phosphorous and potassium detection. Section III discusses the results obtained for each nutrient and lastly section IV concludes the observational study.

2. Proposed Methodology and Implementation

This section describes about the implementation carried out to detect the three essential macro nutrients in plants such as Nitrogen, Phosphorous and Potassium. Two techniques were implemented for stress measurement of different nutrients.

1. Non invasive technique for detection of nitrogen and phosphorous.

2. Invasive technique for detection of potassium nutrient.

Intermediate sub models were built to reach the set goals. Different methods to improve the efficiency by the means of which precision and smart fertigation is obtained is discussed. This part of the paper also confers about different system models that are developed in order to achieve the set goals. It also discusses various methodologies of optimizations to the built models that are adopted to lock the constraints on various criteria's and approach by which appropriate nutrient distribution is carried out.

2.1. Nitrogen Detection

This segment of paper describes the details of nitrogen detection in maize leaf using non-invasive technique by discussing the system architecture proposed and implemented, formulating a mathematical model and optimization. Figure 4 shows the structural view of estimating the content of nitrogen in maize plant. It consists of various sub blocks such as color sensor, Pulse modulation wave processing, ADC block, and comparator and display device. The color sensor was used to sense the color of the leaf. Depending upon the color the amount of nitrogen will be estimated. Constant light source was provided to the leaf. The sensor output will be in digital domain. The color of the leaf changes accordingly with the frequency of the output wave. Visible light and infra red are the two light sources which were considered for experimentation. Different set of readings were obtained depending on the lighting conditions. The digital output obtained from the color sensor is provided from the controller for the signal to be processed. The frequency, wavelength of light reflected from the processed data was calculated with the help of algorithm. The frequency obtained is directly proportional to the reflectance of light that in turn changes with the content of chlorophyll of the leaf. Different frequency values are obtained as per the change in the color of the leaf. The relationship between obtained and reference frequency is gained by using mathematical model.

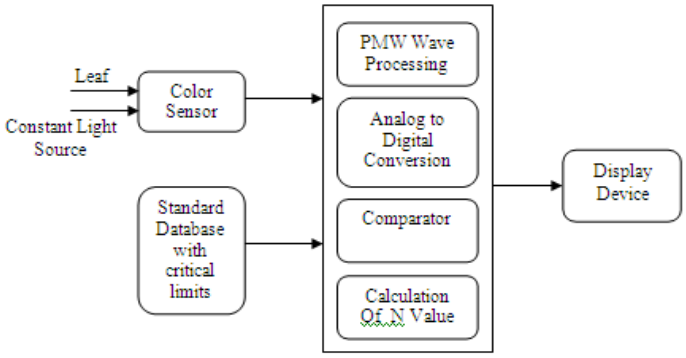


Figure 4: White box of N detecting sub module

The data set for the worst value was aggregated for analytical measurement from UASD College. Based on these values and also the value obtained from the system were compared and the assessed value of nitrogen in percentage of the leaf is unveiled and authenticated.

2.1.1. Mathematical Modeling

There are several authoritative systems in the market which act as chlorophyll meters. SPAD 502 meter is one such device that is made use in this project to act as a reference for spectral assessment of the leaf [13], [14]. Figure 6 provides an insight of the methodology undertaken to arrive at the inter-relationship between the standard device and the device under design.

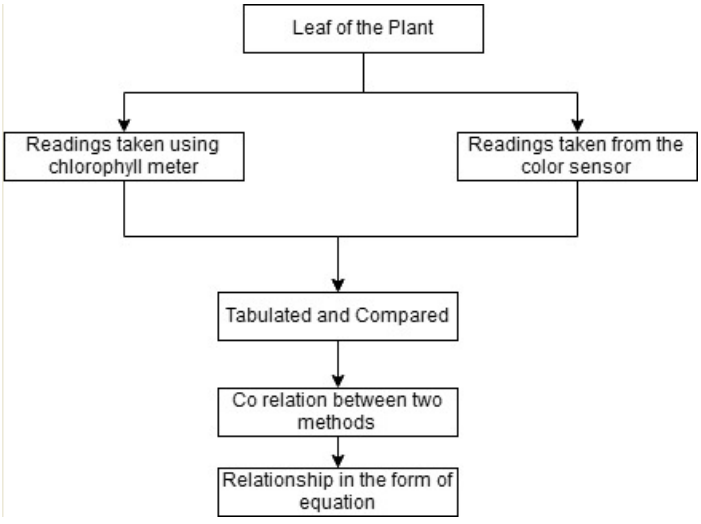


Figure 5: Method to arrive at the relationship between SPAD meter and color sensor

The readings of the leaf using both the devices were taken under different environmental conditions.

Equation (1) describes a linear inter-relation between the data from meter and the color sensor observed by plotting the graph.

$$Sensor = a * SPAD + b \quad (1)$$

Parameter 'a' & 'b' in Equation 1 constitutes to the slope and intercept respectively. This produces a dedicated mathematical model for nitrogen estimation. While creating a data from the reference meter, the readings were taken from all the different areas of the leaf at various locations. The same technique was also carried out while collecting the leaf data using color sensor

The data tabulation of SPAD meter and N_Sensor plot characterizes non- uniformity and was proportional in an inverted manner with each other revealing no patterns. An in-built function in MatLab i.e. fit function was exploited to fit the curve obtained with the help of polynomial of degree. The equation procured after the fitting procedure is as depicted in Equation 2 which shows inverse proportionality relation between the two data sets. Later stage involved modification of Equation 2 in terms of N_Sensor as exemplified in Equation 3.

$$SPAD = (-1.7393)N\_sensor 93.83 \quad (2)$$

$$N\_sensor = (-0.5748)SPAD + 93.83 \quad (3)$$

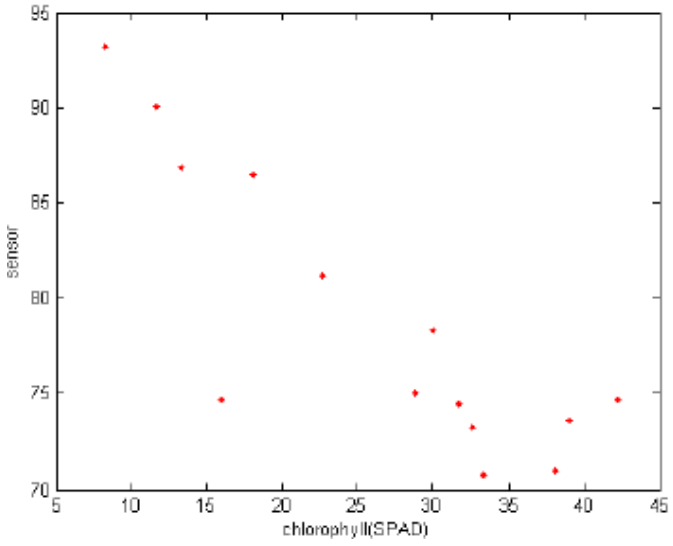


Figure 6: Relationship between chlorophyll content in SPAD and chlorophyll content in sensor

2.1.1. Optimization

Taguchi's technique acts as a reference and narrates a systematic approach for planning and conduction of experiments most effectively when there are numerous parameters related to the end results [15]. Hence the most appropriate data is collected which provides valid and objective conclusions as shown in Figure 7. Figure 8 depicts the inclusion of this approach to obtain different features.

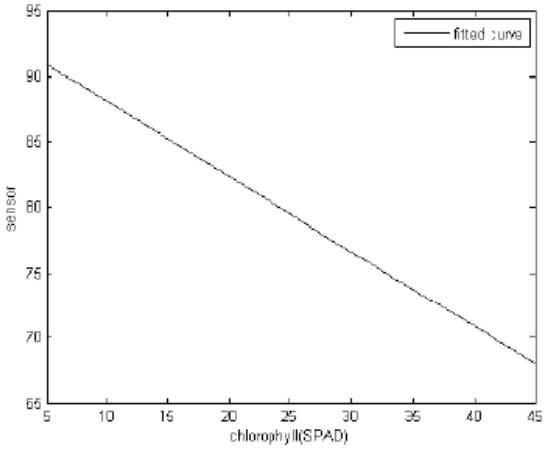


Figure 7: Best fit curve displaying the connection between SPAD device and chlorophyll

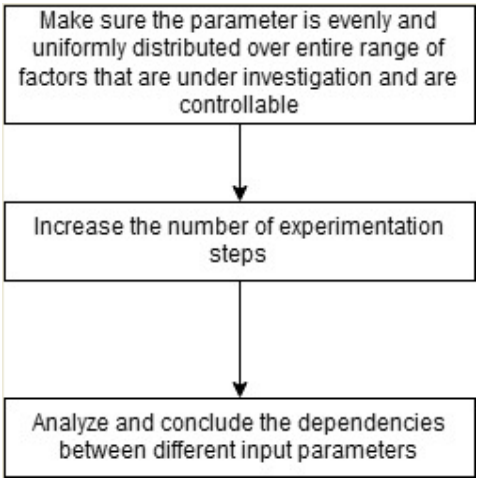
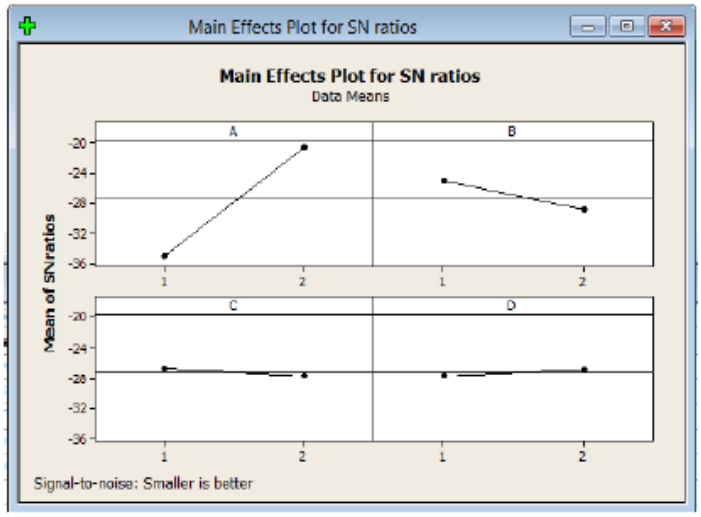
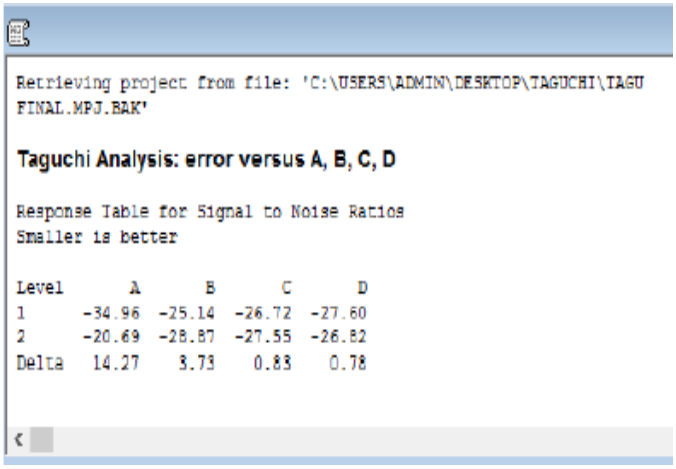


Figure 8: Features list requirement

The signal to noise ratio was plotted and the sensor error was defined as shown in below Figure 9.



In terms of achievement a lower range of the ratio of signal and noise should be present. Hence the gilt edge level of data set should be chosen in that range which has very least level of ratio of noise and signal. After above mentioned experimentation the best result is achieved at 5Volts present at second level with the distance of 1cm and at the lower range of frequency of first level. This also covered a large area at 2nd level as shown in Figure 10. A thorough study was investigated to identify the major design factors affecting the quality of the results using ANOVA technique. The outcome of this investigation led to the conclusion that voltage variation mainly influences the results compared to other design parameters.



2.2. Phosphorous Detection

This is another essential nutrient required in large quantity for the yield to enhance their productivity [16]. As per the observation, this nutrient has absorbance peaks in red regions with the wavelength ranging from 600nm to 700nm. It has no transmission in the near infra-red region. Hence the design in selection of the sensor is enabled such that the emission in light in preferably in the visible region. At the later stage of the sub system an optical filter id designed based on the wavelengths of interest that has to be passed. For this work the wavelength taken into account are beyond 665nm.

To calculate the nutrient amount approximately in the maize, crop a conceptual model was developed as depicted in Figure 11. The instrument used to calculate the phosphorous amount is equipped with series of electro-luminescent diodes that eradicates light with greater intensity in the visible range. The output of the designed system produces the range of phosphorous content in three levels – High, Medium and Low. The specimen of interest is exactly positioned between transmitter and receiver. Later the optical sensor detects the reflectance of those wavelengths.

Table 1: Set Threshold Values on Experimental Basis

Threshold Value	Nature of Leaf
2	No Green
3	Green

Table 2: Readings Obtained for Phosphorous

Readings	Approximate Phosphorous Content
> 3.4	Sufficient
< 3	Deficient

Table 1 discusses about the threshold value set based on the numerous readings that differentiates between green and non-green leaf. Table 2 discusses about the obtained readings which provides an approximate phosphorous content. The advantage of the device under investigation is, it identifies the change in Phosphorous content in an approximate manner well before it is visible by naked eye.

2.3. Potassium Detection

This nutrient is considered as the quality nutrient which is required in large amounts for proper growth as well as reproduction of the plants. The detailed block diagram for detecting the potassium content in coconut leaf is shown in Figure 12.

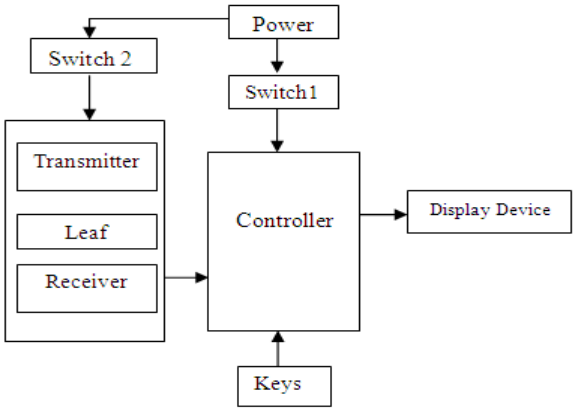


Figure 11: Functional Block Diagram

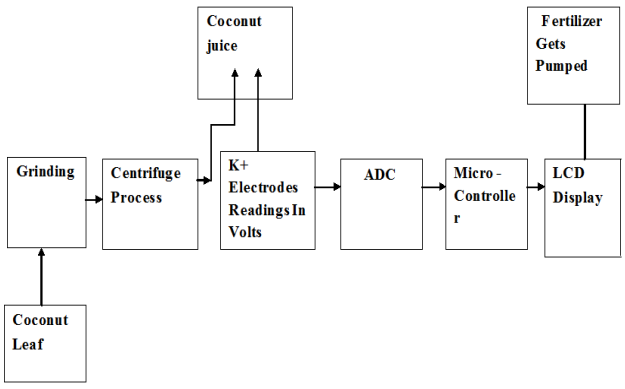


Figure 12: Functional Block Diagram

An experiment was conducted on coconut leaf to estimate the potassium content. One of the invasive techniques, ion selective electrode method was carried out for estimating the nutrient content. In this approach a K^+ electrode is used and the approximate amount of K^+ in a standard unit of PPM is determined.

Table 3: Source- Visible light; Measuring parameter- Nitrogen in maize leaf

Case	Gap between leaf and measuring device in cms	Dark green	Light green	Yellow
Day	2.4	47.05	48.81	53.26
	2.7	45.33	47.07	51.81
	3	43.10	46.01	50.95
Night	2.4	46.75	46.18	48.16
	2.7	44.83	45.98	46.445
	3	43.46	44.48	45.66

Table 4: Source-Infrared; Measuring parameter- Nitrogen in maize leaf

Case	Gap between leaf and measuring device in cms	Dark green	Light green	Yellow
Day	2.4	46.23	54.56	62.05
	2.7	37.27	46.34	52.34
	3	31.07	39.8	43.07
Night	2.4	41.3	45	50.06
	2.7	36.76	39.09	41.89
	3	30.12	32.76	36.23

3. Results and Discussion

This part of the study details about the tabulation of results obtained from the designed system for nutrient management in plants. Initial sub sections discuss the detection of nitrogen and phosphorous content in maize plant using non-invasive technique and the last sub section describes the results obtained for potassium content in coconut leaf using invasive technique. An IoT application was built to store the acquired data onto the cloud for further analysis. The created home page with experimental values is been listed.

3.1. Nitrogen Detection

SPAD meter is used as a reference device to carry out the investigation work. The data was recorded and the average of sensor data was calculated for each segment of the leaf. There is a need to compare a wide range of chlorophyll data which is obtained from the spectroscopy. The readings were considered from different segments of the leaf. The most important observation was that the content of chlorophyll was decreased at the center portion of the maize leaf. As it is apparent that leaf's nitrogen content directly varies with the content of chlorophyll in it. Thus, there is a linear relation between the two. Exploiting this mathematical relation maize leaf nitrogen content is calculated. Tests are conducted during day and night for validating the impact of sunlight on the designed sensor. The sensor is also validated for different sources of light particularly visible and infrared light as depicted in Table 3 and 4

T

Figure 13: Prototype showing high nitrogen in maize leaf

Figure 13 exhibits the proof of concept of the sub-component implemented.

3.2. Phosphorous Detection

For proof of concept the specimen leaf of maize plant was taken into account. A total of 10 leaves were procured on random basis for analysis. The test was conducted on the leaf using two methods for validation. The work carried out mainly concentrates on developing a strong interdependency between nutrient levels of phosphorous and reflectance numbers in two regions i.e. red and NIR range. [17].

Figure 14 & 15 displays the values obtained by the component under design and web page that includes the real time readings.

Figure 14: Home Page

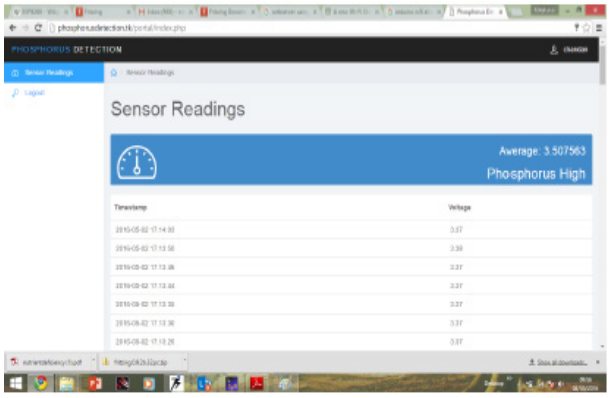


Figure 15: Dash Board

3.3. Potassium Detection

The measuring device initially requires preparation of coconut leaf extract that involves grinding followed by centrifuge process. The electrodes are used to estimate the reading in volts and the data is refined using signal conditioning circuits. The controller processes the conditioned data and displays the results of the nutrient content in terms of High (H), Medium (M) and Low (L). Table 5 displays the results obtained by ion selective method for potassium content in the coconut leaf. There was a linear relationship between the electrode output readings and concentration in PPM. This method of detecting the potassium content was further corroborated by taking the coconut leaf samples for UAS Dharwad which was categorized into less potassium content, medium content and rich in potassium. Using these leaves the results were obtained that stated the potassium content as per its presence in the leaf.

Table 5 : Results Obtained for Potassium Nutrient

Electrode Output in Volts	Concentration in PPM
0.2	4.503
0.4	20.28
0.6	106.204
0.8	478.3
1.0	2154.43
1.2	9703.49

4. Conclusion

Irrigation and fertigation play a pivotal role in better yield of crops. Excessive usage of fertilizers, in particular nitrogen leads to evaporation of this chemical which in turn creates global warming. Hence this impacts the environment. The projected work assesses the quantity of the nutrient present in the leaf approximately in the scale of high, medium and low. Three nutrients are addressed which are treated as macro nutrients. The nitrogen and phosphorous was tested and validated for maize plant where as potassium nutrient was initially experimented from coconut leaf. The entire work is based on the spectral analysis which was conducted in isolation for N and P and by inculcating non-invasive method of stress detection. Compared to conventional procedure, sensor-based reasoning of nutrients contributes to large spectrum of readings. Different optimization techniques were involved and among them Taguchi's approach was used in nitrogen detection.

A server is created to observe the real time data remotely. An invasive technique was also subjected in the proposed work for detecting potassium content in coconut leaf by using ion selective electrode method which was found to be time consuming and involving huge devices for making the required solution.

5. References

[1] B. H. Shraddha, R. M. Shet, P. Nikita and N. C. Iyer, "Nutrient Detection for Maize Plant Using Noninvasive Technique," *2018 International Conference On Advances in Communication and Computing Technology (ICACCT)*, Sangamner, 2018, pp. 504-509.doi: 10.1109/ICACCT.2018.8529659

[2] Ghorbani, Amir Mirzaei Mossivand and Abazar Esmali , "Utility of the Normalised Difference Vegetation Index (NDVI) for land/canopy cover mapping in Khalkhal County".Ouri 2012

[3] Bodo M, Schmidhalter U ," Estimating the nitrogen nutrition index using spectral canopy reflectance measurements". Europ J Agron. 29:184-190.,2008.

[4] Betula pendula Roth,"Effects of nitrogen on growth properties and phenology of silver birch" JOURNAL OF FOREST SCIENCE, 58, 2012 (9): 391–399

[5] Guijun Yang; Chunjiang Zhao; Ruiliang Pu; Haikuan Feng; Zhenhai Li; Heli Li; Chenhong Sun; "Leaf nitrogen spectral reflectance model of winter wheat"(Triticum aestivum L).

[6] J.E. McMurtrey ; E.M. Middleton ; L.A. Corp ; P.K.E. Campbell ; L.M. Butcher ; C.S.T. Daughtry "Optical reflectance and fluorescence for detecting nitrogen needs in Zea mays L."

[7] Van Den Berg, A.K.; Perkins, T.D. Evaluation of a portable chlorophyll meter to estimate chlorophyll and nitrogen contents in sugar maple (Acer saccharum Marsh.) leaves. For. Ecol. Manag. 2004, 200, 113–117.

[8] P Nikita, P Jyoti, B Shraddha, P Preeti, S Raghavendra, Nalini C Iyer, " Active Learning in Electronic Measurements and Instrumentation Course through Hands-on " Journal of Engineering Education Transformations.

[9] Ahmad U, Mardison S, Tjahjohutomo R, Nurhasanah A , " Developmentof automatic grading machine prototype for citrus using imageprocessing," Aust J Agric Eng. 1(5):165-169. , 2010.

[10] BH Shraddha, M Subhas, P Nikita, P Preeti, K Bhagyashree, RM Shet, P Jyoti, Nalini C Iyer, "Model based Learning of Linear Integrated Circuit Course: A practical approach" Journal of Engineering Education Transformations.

[11] Anatoly A. Gitelson, Yuri Gritz and Mark N. Merzlyak "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves".

[12] Anand M, Haripriy M, Byju G, "Chlorophyll meter and leaf colour chart to estimate chlorophyll content, leaf color, and yield of cassava.Photosynthetic",. 46: 511-516., 2008.

[13] M. T., Voroney, P. and Parkin, G. 2005, "Predicting nitrogen fertilizer requirements for corn by chlorophyll meter under different N availability conditions". Can. J. Soil Sci. 85: 149159.

[14] Chang SX, Robison DJ ," Nondestructive and rapid estimation of hardwood foliar nitrogen status using the SPAD-502 chlorophyll meter". Forest Ecol Manag. 181: 331-338.,2003.

[15] Preeti Pillai, S Raghavendra, B Shraddha, P Nikita, Nalini Iyer, " Digital Signal Processing: An abstract mathematics to real world experience" Journal of Engineering Education Transformations.

[16] Osborne, S. L., J. S. Schepers, D. D. Francis, and M. R. Schlemmer. 2002. "Detection of Phosphorus and Nitrogen Deficiencies in Corn Using Spectral Radiance Measurements".

[17] Ciganda V, Gitelson A, schepers J," Non-destructive determination of maize leaf and canopy chlorophyll content. J Plant Physiol" 166: 157-167.,2009.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 135-141 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Design and Implementation of State-PID Feedback Controller for Poultry House System: Application for Winter Climate

Ilyas Lahlouh*, Ahmed Elakkary, Nacer Sefiani

Mohammed V University of Rabat, High School of Technology of Sale (EST), LASTIMI Laboratory, System Analysis, Information Processing and Industrial Management, 11030, Morocco

ARTICLE INFO

Article history:
Received: 08 September, 2019
Accepted: 04 January, 2020
Online: 22 January, 2020

Keywords:
State-PID feedback controller
Poultry house
Disturbance
Stability analysis

ABSTRACT

The poultry house is the area where the chickens are maintained for the main purpose to improve the productivity and the environmental conditions for the broilers. The growth of chickens inside the poultry house can be affected by several criterions such as relative humidity and temperature. In this paper, we propose to design and implement experimentally a state-PID feedback controller in order to achieve a high stabilization of the dynamics systems of the poultry process. The purpose of this study is to keep the temperature and relative humidity at desired values and to eliminate the disturbance generated during the winter climate. The effectiveness of the proposed controller is evaluated through numerical examples and with an experimental poultry house prototype. To feature more the efficiency of the proposed approach implemented, a comparative study has been conducted between the results achieved by the proposed method and the state-PI feedback controller and that of a state feedback controller.

1. Introduction

The poultry house climate is defined as the principal part that can straightforwardly influence the efficiency and comfort of the raising birds. Notwithstanding, the control of the housed livestock building remains a confounded errand because of the nonlinear behavior of psychrometric mechanisms required on both physiological and psychological components.

In most cases, the poultry farming is mainly affected by the outside condition (external relative humidity and temperature) and by the control of the actuators of the system (PAD cooling, heating, and ventilation system) [1]. Recently, the agricultural sector has extended altogether regarding control and optimization strategies for the primary parameters influencing good livestock management.

The hygro-thermal parameters are the most critical environmental variable on the poultry as announced in [2], [3]. Additionally, it has mutually been shown that the relative humidity and temperature affect intently the animal's improvement in Morocco because of the dominating weather and the country's location.

In this regard, we target our works to extend the ongoing studies by designing a suitable control strategy that is created depending on the mathematical model of broiler houses under cold conditions.

Over the most recent couple of years, the late research investigation demonstrates a tendency to the utilization of progressively different advanced procedures for control systems and simulation for a livestock building.

For the most part, it exists two types of control techniques, traditional techniques, and modern techniques. In such manner, numerous researches on livestock building concerning climate-controlling have been suggested in the literature, for example, nonlinear robust control for pig house [4], nonlinear adaptive control for animals house [5], predictive control for temperature and humidity in a normally ventilated structure [6], SMC (Sliding Mode Control) to control both of interior temperature and relative humidity in an evaporative cooling system [7].

Others studies related to the optimization control have been developed in the livestock building, for instance, the application of

*Corresponding author: Ilyas Lahlouh, High School of Technology of Sale, +212676640500, Email: ilyaslahlouh@research.emi.ac.ma

the ant colony optimization (ACO) for control broiler house under hot climates [8] and the utilization of the particle swarm optimization (PSO) in the managing of the thermal requirement of the poultry house.

In other hands, the fuzzy logic controller has been well tested in a greenhouse and poultry house [9], authors are reasoned that is fuzzy controller was able to give a superior efficiency in the stability. Other researchers [10] and [11] have linked a PID with fuzzy controllers to control the temperature indoors poultry house, they had concluded that the fuzzy-PID structure has excellent performance over a PID controller as far as settling time and steady-state error. A comparison between fuzzy and on/ off controllers have been analyzed in [12] to test and keep up the ambient parameters inside a model poultry house. Accordingly, all these advanced control systems have demonstrated their similarity to be used on control structure applications for poultry houses.

In this paper, the present work describes an extension of the original work reported in Proceedings of the International Conference on Automation and Computing [13].

The contributions of this article are portrayed as looks for: (1) to design a state-PID feedback controller to guarantee nominal stability and to regulate the hygro-thermal parameter of the system to a given set point value. (2) Implement the proposed controller in a real poultry house prototype system.

Hence, the inspiration for the state-PID feedback controller in this manuscript originates from some papers dealing with a wide variety of systems such as the SISO and MIMO systems [14] ,[15],[16].

The control configuration begins with the deduction of the linearized hygro-thermal model of the poultry process proposed in [17], at that point develops the state-PID feedback controller utilizing the eigenvalue assignment algorithm.

The remainder of the paper is formulated as follows: the primary segment exhibits quickly the hygro-thermal model of the poultry process. The second section describes the methodology of the state-PID feedback controller (SFCPID). Finally, some simulation results and remarks for stabilization and disturbance rejection are given to show the effectiveness of the proposed controller.

2. Mathematical model

2.1. Hygro-thermal model

Considering the mathematical modeling of the previous work detailed in [17] and the proposed studies on climate modeling inside livestock buildings [6] [18], the hygro-thermal model administered the poultry house is:

$$\dot{x_1} = \frac{u_1}{a} - \frac{r}{a}u_2 + R_{cb}(T_b - x_1) - (\frac{K_g}{a} + \frac{u_3}{V_b})(x_1 - T_{ext}) \quad (1)$$

$$\dot{x_2} = \frac{u_2}{b} + R_{ev}(0,26x_1^2 - 6,46x_1 + 81,6) - \frac{u_3}{V_e}(x_2 - H_{ext})$$

Where:

$$a = \rho_{air} V_b C_p \qquad b = \rho_{air} V_e \qquad R_{ev} = \frac{N_c.0,001}{3600.b}$$

$$R_{cb} = \frac{N_c.0,081.M_i^{0,67}}{\rho_{air} V_b C_p (\overline{R_{coat(i)}} + \frac{1}{5,4 + \frac{0,052 + 1,55.\vartheta_i^{0,48}.M_i^{0,16}}{0,13.M_i^{0,33}}})}$$

The state space form of the model (1) is represented in equation (2). The study is conducted with the cold operations where the controller input related to the PAD cooling system isn't working ($E_{ev} = U_2 = 0$).

$$\begin{bmatrix} \dot{\delta x_1} \\ \dot{\delta x_2} \end{bmatrix} = \begin{bmatrix} -(R_{cb} + \frac{K_g}{a} + \frac{\overline{u_{3e}}}{V_b}) & 0 \\ R_{ev}(0,52\overline{x_{1e}} - 6,46) & -\frac{\overline{u_{3e}}}{V_e} \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix} + \begin{bmatrix} \frac{1}{a} & -\frac{\overline{x_{1e}}}{V_b} \\ 0 & -\frac{\overline{x_{2e}}}{V_e} \end{bmatrix} \begin{bmatrix} \delta u_1 \\ \delta u_3 \end{bmatrix} + \begin{bmatrix} \frac{K_g}{a} + \frac{\overline{u_{3e}}}{V_b} & 0 \\ 0 & \frac{\overline{u_{3e}}}{V_e} \end{bmatrix} \begin{bmatrix} \delta d_1 \\ \delta d_2 \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} \delta Y_1 \\ \delta Y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix}$$

Where x are states variables associated with the internal relative humidity and temperature, u_1 and u_2 are the input controller associated with the heating system and the ventilation system, d are external disturbances.

The linearized model (2) can be simplified and represented to as a state space structure yielded (3):

$$\begin{aligned} \dot{\delta}_x &= A\delta_x + B\delta_u + B_d\delta_d \\ \delta\dot{y} &= C\delta x \end{aligned} \quad (3)$$

Where:
C is the output matrix, B and B_d are the input matrix and A is the state matrix.

2.2. Local Stability

We recommended concentrating to study the global stability of the general system (nonlinear) from its linearized tangent.

Let us consider the nonlinear system expressed as:

$$\begin{aligned} \dot{x} &= f(x) \\ f(x_e) &= 0 \end{aligned} \quad (4)$$

The linearized of the system around the operating point x_e is:

$$\delta\dot{x} = A(x_e)\delta x$$
$$\delta x = x - x_e$$

$$A(x_e) = \left(\frac{\partial f}{\partial x}\right)_{x_e} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & .... & \frac{\partial f_1}{\partial x_n} \\ . & ..... & . \\ . & & . \\ \frac{\partial f_n}{\partial x_1} & .... & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

After calculating, we get:

$$A(\overline{x_{1e}}) = \begin{bmatrix} -(R_{cb} + \frac{K_g}{a} + \frac{\overline{u_{3e}}}{V_b}) & 0 \\ R_{ev}(0{,}52\overline{x_{1e}} - 6{,}46) & -\frac{\overline{u_{3e}}}{V_b} \end{bmatrix}$$

If the case of $\overline{u_{3e}} = D_v = 0$, especially when the livestock building system work just with the heating system ($Q_T \neq 0$).

The eigenvalues of the matrix are:

$$\lambda_1 = 0 \qquad \text{and} \qquad \lambda_2 = -(R_{cb} + \frac{K_g}{a})$$

It is clear that the process is locally on the cutoff of the stability, and requires stabilization with an appropriate controller.

3. State-PID feedback controller design

In this section, state feedback and PID controllers are designed for the linear state space proposed model of the broiler house process.

3.1. State feedback controller

Let's consider the LTI system which having the following form:

$$\dot{x} = Ax + Bu \tag{5}$$

Where:

$A \in R^{nxn}$ $B \in R^{nxm}$ $x \in R^{nx1}$ $u \in R^{mx1}$

The linear system under consideration must to be totally controllable, the controllability matrix must have a full rank-n [19][20] and can be communicated as

$$C = [B \quad AB \quad A^2B \quad .... \quad A^{n-1}B] \tag{6}$$

The pair (A, B) is controllable.

Figure 1 illustrates the basic feedback controller of the poultry house.

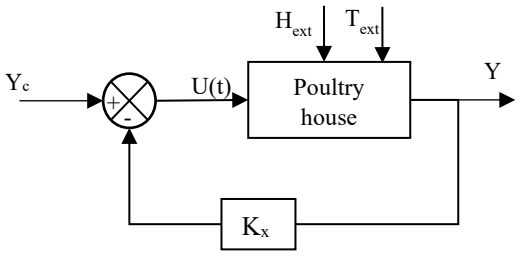


Figure 1: State feedback controlled poultry house system.

Where Y_c, U, Y and K_x are the command signal, the control law, the output of the plant and the feedback controller, respectively.

The control law can be composed on the structure:

$$u = Y_c - K_x x \tag{7}$$

The objective of this controller is to stabilize the system by the constant feedback control that imposes to place the poles at desired characteristic behavior for the states.

Using the feedback control law (7), the closed-loop characteristic polynomial system is given by:

$$|sI - (A - BK_x)| = 0 \tag{8}$$

The structure issue is to determine the feedback gain matrix K_x such that the closed-loop poles $\{\lambda_1, \lambda_2, ...\lambda_n\}$ of the system (8) are appointed at the desired values.

The methodology proposed to this problem is accomplished by using the pole placement algorithm.

The assignment placement algorithm concedes finding the matrix K_x, such as the closed-loop, system is stable. Both λ_i the desired closed loop poles and ν_i the eigenvector relating.

The following steps present the algorithm used for the assurance of the gain components of matrix K_x is: [21]

1. $(A - BK_xC)\nu_i = \lambda_i \nu_i$ Where : $K_x \in R^{mxn}$
2. $\nu_i = -(\lambda_i I - A)^{-1} BK_xC\nu_i$
3. $L_i = -(\nu_i I - A)^{-1} B$ With : $L_i \in R^{nxm}$

   $z_i = K_xC\nu_i$ With: $z_i \in R^m$
4. $z_i = (L_i^T R_i L_i)^{-1} L_i^T R_i \nu_i^d \quad i = 1,....,m$
5. $\nu_i^a = L_i z_i \qquad i = 1,....,m$

   $J_i = \frac{1}{2}(\nu_i^d - L_i z_i)^T R_i (\nu_i^d - L_i z_i)$
6. $K_x = ZV_a^{-1}$

With: $Z = [z_1 \quad z_2 \quad ... \quad z_m]$ and

$V_a = [\nu_1^a \quad \nu_2^a \quad ... \quad \nu_m^a]$

3.2. State-PID feedback controller

In this area, we introduce the proposed strategy control; we use in conjunction the state feedback and the PID controller.

The Proportional Integrator Derivative structure is one of the most controller broadly utilized in the industry, covering more than 90% of the industrial need in control. This predominance is because of its simplicity to implement it in a large plant does not require detailed knowledge. The control block scheme is shown in figure 2.

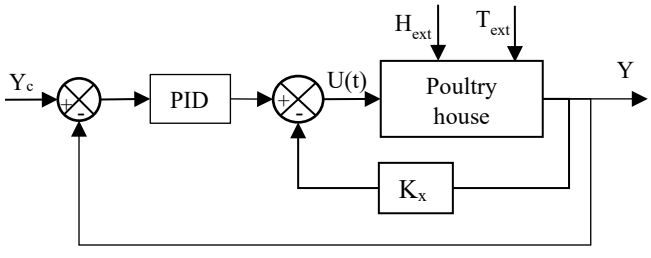


Figure 2: State-PID feedback proposed controller.

The control law u of the state-PID feedback proposed controller is :

$$U = K_p e + K_i \int e(t)dt + K_d \frac{de(t)}{dt} - K_x x \qquad (9)$$

Where $(K_p, K_i, K_d) \in R^{nxm}$ are the planned gain matrices to achieve a desired closed-loop system.

In summary, the problem described is resolved for a pole placement method consisting of three separated steps as mentioned in Guo et al [22].

The pole placement by state-P feedback is directed first prompting an intermediate system. Furthermore, state-I feedback is executed; another transitional system resulted. At last, the closed-loop system with the desired characteristic polynomial is acknowledged by state-D feedback joined as the final plan configuration.

The components of state-PID controller gain matrices for the PID structure can be obtained using the transformation into Frobenius canonical form detailed in [23],[24] and the equations mentioned in [15].

4. Results and discussions

In this analysis, we suggest regulating the internal temperature T_{int} and relative humidity H_{int} under cold climates, for that we design the state-PID feedback controller to get synchronization to deal with both ventilation and heating systems.

To affirm the performance of the proposed state-PID feedback controller, several comparative illustrative simulations are delineated.

4.1. Case 1: simulation study

The industrial poultry house system was equipped with controller natural ventilation, heating and PAD cooling system, the livestock building dimensions were 12,4m x 120m x 3,85m.

The real plant data have been saved for the entire period for the chickens raising of [20-21] days under the climates of the Mediterranean region (Rabat area, Morocco). (Figure 3)

Figure 3: Poultry house system used in the case study.

Table 1 presents the numerical estimations of the distinct criterions used in the simulation of the proposed controller.

Table 1: Input specifications used in the hygro-thermal model

Specifications	Values	Specifications	Values
ρ_{air}	1,2kg/m^3	T_b	41°C
C_p	1006J/kg.K	N_c	24000
V_b	3033m^3	V_e	3033m^3
M	0,905kg	ϑ	0,2m/s
$\overline{R_{coat}}$	0,2 m^{2}K/W	K_g	3126W/°C

Design of state feedback controller

The conventional (SFC) design technique begins with the determination of the desired eigenvalues from the given specifications of the controller K_x : λ_1^d , λ_2^d of an order of multiplicity respectively (n) and (n-1).

The basis of controllability is verified of the system (2) after including the different parameters:

Rank (C) =2

The specifications considered for the poultry house system are selected primarily on diminishing the steady-state error and the settling time.

After choosing the desired pole and applying the assignment eigenvalues algorithm, we find:

$$K_x = 10^3 \begin{bmatrix} 3659 & -51282 \\ -0 & -2 \end{bmatrix}$$

The input control of the system is :

$$u = 10^3 \begin{bmatrix} -3659 & 51282 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} T_{int} \\ H_{int} \end{bmatrix}$$

Design of state-PID feedback controller

The feedback gain matrix is computed as:

$$K_x = 10^3 \begin{bmatrix} 25637 & -76923 \\ -0 & -3 \end{bmatrix}$$

The PID gain matrices is obtained as :

$$K_p = 10^3 \begin{bmatrix} 3663 & -51282 \\ -0 & -2 \end{bmatrix} \qquad K_i = 10^3 \begin{bmatrix} 43956 & -51282 \\ -0 & -2 \end{bmatrix}$$

$$K_d = 10^3 \begin{bmatrix} 25637 & -76923 \\ -0 & -3 \end{bmatrix}$$

Stability analysis

The comparison of the closed loop response of the hygro-thermal parameters concerning the desired values of poultry house system with conventional state feedback, state-PI feedback and state-PID feedback controller using the package Simulink of MATLAB are demonstrated in figure 4 and 5.

For numerical simulation, the desired relative humidity and internal temperature have been fixed at 66% and 26°C.

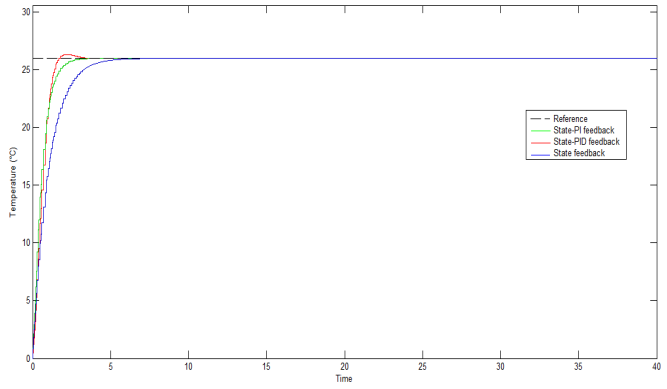


Figure 4: Temperature response generated by the different controller.

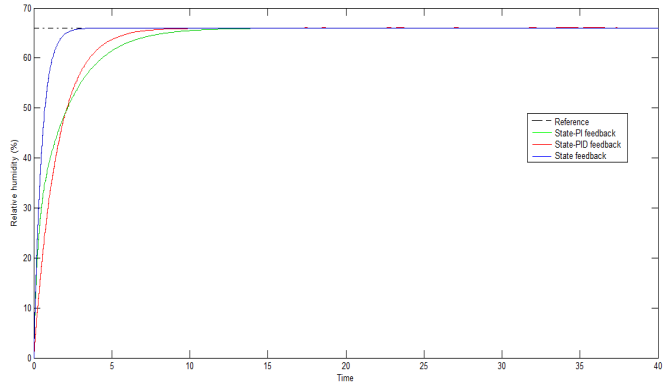


Figure 5: Humidity response supplied by the different controller

The state-PID feedback controller configuration performed wells; however, the state feedback controller performs moderately. It can be clearly seen in figure 4 and 5 that there are different responses in the transient state performance. The numerical values of these performances of the closed loops are tabulated in table 2 for comparison.

Table 2: Numerical values of standard performance indices.

	Performances indices	SFC	SFCPI	SFCPID
Temperature	Rise time (s)	2,42	1,38	1,21
	Settling time (s)	3,128	1,66	1,49
	Overshoot (%)	0	0	1,54
Relative humidity (%)	Rise time (s)	1,12	4,14	3,76
	Settling time (s)	1,31	5,62	4,37

From table 2, it appears that the state feedback controller gives poor results compared to that of state-PI and PID feedback controller in control of temperature. Other hands, the SFCPI gives a slow response compared to that of state-PID feedback state and a state feedback controller. Furthermore, the SFCPID affords the littlest value of rise time evaluated at 1,21s and the smallest value of settling time estimated at 1,49s.

It tends to be additionally observed that the performances indicated by the proposed controller are closely at their best values. The transient responses of states are plotted in figure 6 and 7.

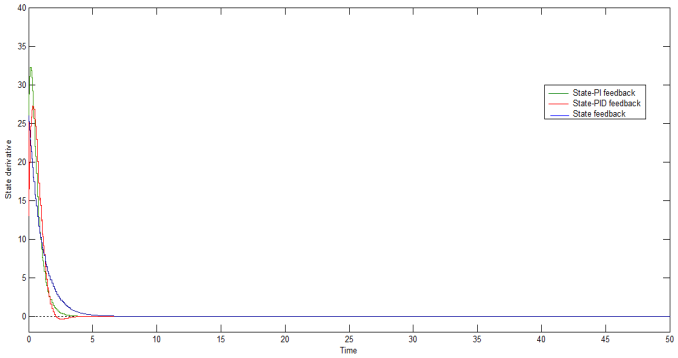


Figure 6: Response of the state derivative of the internal temperature.

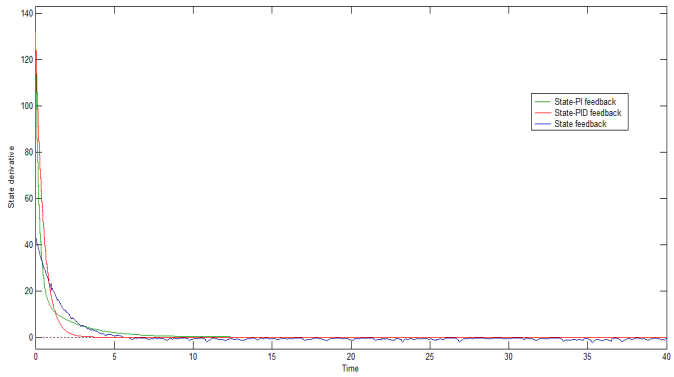


Figure 7: Response of the state derivative of the relative humidity.

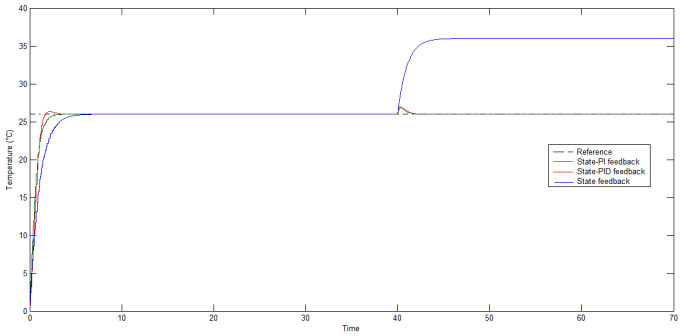


Figure 8: Temperature response simulated under disturbance with the different controller.

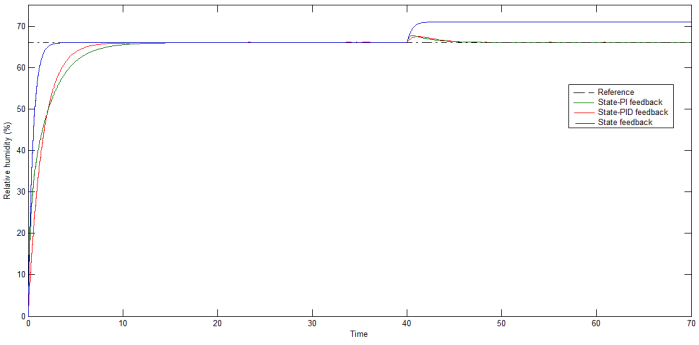


Figure 9: Relative humidity response simulated under disturbance with the different controller.

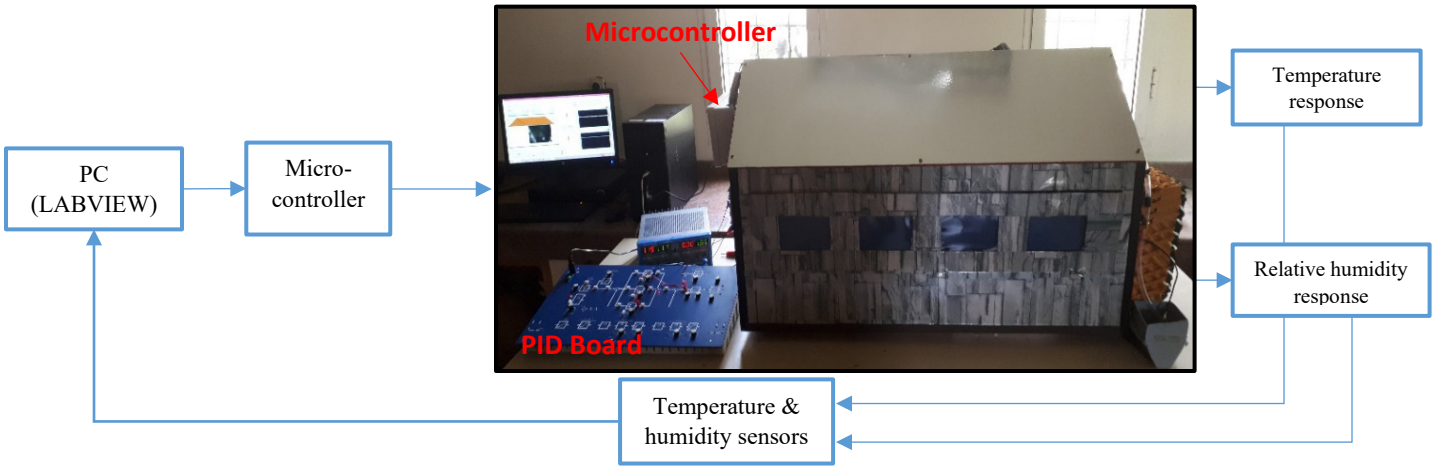


Figure 10: Experimental poultry house prototype.

From figures 6 and 7, it can be noticed that the derivative of the internal temperature accelerates rapidly within the state-PID feedback controller compared to that of SFCPI and that of state feedback. Also, it can be observed that the derivative response converges to zero after a short duration and this proves that the internal temperature and relative humidity have been well stabilized.

Disturbance rejection

The temperature and humidity response under the external disturbance with the three controllers are displayed in figures 8 and 9.

From figures 8 and 9, it very well may be obviously observed that with our proposed controller that the system reactions quicker, and recuperates totally from the external disturbance without any residue errors. This is because of a derivative and integral parameters acting as an error corrector.

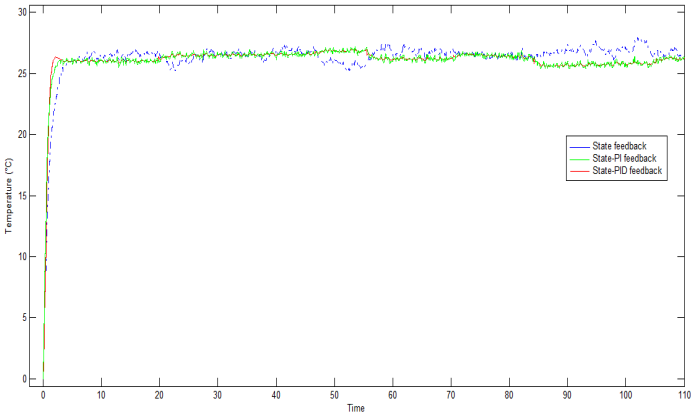
Furthermore, it tends to be seen that the planned state-PID feedback controller maintains the internal temperature and relative humidity within a short time of 7 seconds after applying a disturbance; however, the conventional state feedback controller isn't able to reduce or eliminate the external disturbance in both of temperature and relative humidity response.

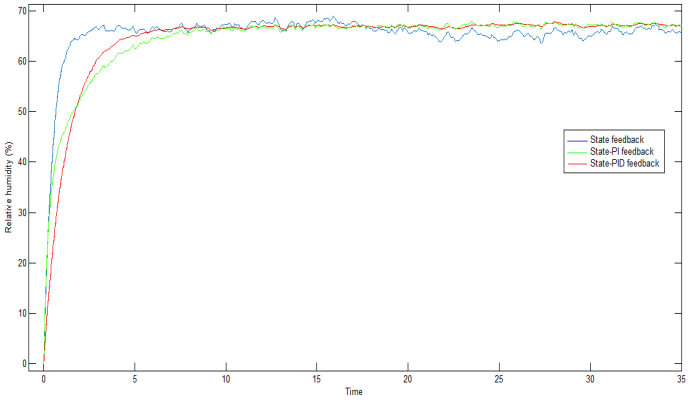
4.2. Case 2: Experimental poultry house prototype

The poultry house prototype has two main actuators, namely, the electric heater and the fans systems. Other subsystems contain lighting and evaporative cooling systems are also installed. The system program includes the main algorithm unit for all systems in the broiler house. In addition, the experimental data is transmitted to the computer device using a micro controller. The real data is monitored with the LabVIEW software; the users can control the important operations and can to be informed of the various changes for the system. (See figure 10)

The internal temperature is measured using LM35 and the relative humidity is acquired via HIH4030 sensor. The variable input of heating and ventilation system are fitted with a thyristor driver circuit in order to produce a varying voltage for running the fans and the heater at different level speed and stage. The change in applied voltages is legitimately relative to the airflow speed and heating rate.

The three controllers are implemented for the poultry house prototype and experimental results of the temperature and relative humidity response are represented in figures 11 and 12.





The implementation of the different controller to the poultry house prototype control system show the better quality of using SFCPID in the regulation of the system. The results obtained with

the state-PID feedback controller provide a better response for relative humidity and temperature response with a very smooth variation in tracking the desired set points.

5. Conclusions

In this article, we have proposed and designed a state-PID feedback controller for stabilizing the relative humidity and temperature indoor the broiler house process during the colds conditions.

Numerical examples are tested to confirm the effectiveness of the proposed design controller to stabilize the system. On the other, we have considered the existence of disturbance simulated as the external perturbations getting from the outside climate. The results of simulations show that the proposed methodology was proficient regarding dynamic performance, for example, a decrease of settling time and rise time. Moreover, the SFCPID was capable of tracking the desired set point with a rapid rate of convergence. Thus, the proposed controller has proven its performance through the simulations under disturbance and the comparative study conducted between the state feedback and state-PI feedback technique and the state-PID feedback controller.

In addition, the proposed controller is implemented for the real experimental poultry house prototype. It is concluded that the proposed state-PID feedback controller affords excellent performance than the state-PI feedback and state feedback controller in terms of following the set points parameters.

Finally, it is pointed out that the proposed control methodology of the poultry house accentuates the regulation of the system and the optimal parameters.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by the National Center for Scientific and Technical Research (CNRST) within the context of the Research Excellence Awards program.

References

[1] S. Yahav, A. Straschnow, D. Luger, D. Shinder, J. Tanny, and S. Cohen, "Ventilation, sensible heat loss, broiler energy, and water balance under harsh environmental conditions," *Poult. Sci.*, vol. 83, no. 2, pp. 253–258, 2004.

[2] N. J. Daghir, Ed., *Poultry production in hot climates*, 2nd ed. Wallingford: CABI, 2008.

[3] T. Banhazi *et al.*, "Issues related to livestock housing under hot climatic conditions including the animals' response to high temperatures," *Anim. Hous. Hot Clim.*, p. 4, 2007.

[4] A. G. Soldatos, K. G. Arvanitis, P. I. Daskalov, G. D. Pasgianos, and N. A. Sigrimis, "Nonlinear robust temperature–humidity control in livestock buildings," *Comput. Electron. Agric.*, vol. 49, no. 3, pp. 357–376, Dec. 2005.

[5] P. I. Daskalov, K. G. Arvanitis, G. D. Pasgianos, and N. A. Sigrimis, "Non-linear Adaptive Temperature and Humidity Control in Animal Buildings," *Biosyst. Eng.*, vol. 93, no. 1, pp. 1–24, Jan. 2006.

[6] P. I. Daskalov, "Prediction of temperature and humidity in a naturally ventilated pig building," *J. Agric. Eng. Res.*, vol. 68, no. 4, pp. 329–339, 1997.

[7] K. Senawong, S. Winitchai, and T. Radpukdee, "Humidity and temperature control in an evaporative cooling system of a poultry house," *Eng. Appl. Sci. Res.*, vol. 39, no. 1, pp. 95–111, 2012.

[8] I. Lahlouh, A. El Akkary, and N. Sefiani, "PID/Multi-Loop Control Strategy for Poultry House System Using Multi-Objective Ant Colony Optimization," *Int. Rev. Autom. Control IREACO*, vol. 11, no. 5, p. 273, Sep. 2018.

[9] K. Chao, R. S. Gates, and N. Sigrimis, "Fuzzy logic controller design for staged heating and ventilating systems," *Trans. ASAE*, vol. 43, no. 6, p. 1885, 2000.

[10] D. O. Aborisade and O. Stephen, "Poultry house temperature control using Fuzzy-PID controller," *IJETT Chennai*, vol. 11, no. 6, pp. 310–314, 2014.

[11] R.Manoj Manjunath and S.Janaki Raman, "Fuzzy adaptive PID for Flow Control System based on OPC." IJCA Special Issue on "Computational Science-New Dimensions & perspectives," 2011.

[12] E. Mirzaee-Ghaleh, M. Omid, A. Keyhani, and M. J. Dalvand, "Comparison of fuzzy and on/off controllers for winter season indoor climate management in a model poultry house," *Comput. Electron. Agric.*, vol. 110, pp. 187–195, Jan. 2015.

[13] I. Lahlouh, A. El Akkary, and N. Sefiani, "State feedback controller in a closed poultry house system," in *24th International Conference on Automation and Computing*, Newcastle upon Tyne, 2018.

[14] W. Wiboonjaroen and S. Sujitjorn, "State-PID Feedback for Magnetic Levitation System," *Adv. Mater. Res.*, vol. 622–623, pp. 1467–1473, Dec. 2012.

[15] S. Sujitjorn and W. Wiboonjaroen, "State-PID Feedback for Pole Placement of LTI Systems," *Math. Probl. Eng.*, vol. 2011, pp. 1–20, 2011.

[16] I. H. Akyuz, E. Yolacan, H. M. Ertunc, and Z. Bingul, "PID and state feedback control of a single-link flexible joint robot manipulator," in *2011 IEEE International Conference on Mechatronics*, 2011, pp. 409–414.

[17] I. Lahlouh, A. El Akkary, and N. Sefiani, "Mathematical Modelling of the Hygro-Thermal Regime of a Poultry Livestock Building: Simulation for Spring Climate," *Int. Rev. Civ. Eng. IRECE*, vol. 9, no. 2, p. 79, Mar. 2018.

[18] International Commission of Agricultural Engineering, Ed., *CIGR handbook of agricultural engineering*. St. Joseph, MI: American Society of Agricultural Engineers, 1999.

[19] G. Conte, C. . Moog, and A. M. Perdon, *Algebraic methods for nonlinear control systems*. London: Springer-Verlag London Limited, 2007.

[20] P. J. Antsaklis and A. N. Michel, *Linear Systems*. Boston (MA): Birkhäuser, 2006.

[21] R. Fonod and P. Kocsis, "State Feedback Control Design Using Eigenstructure Decoupling," in *18th International Conference on Process Control*, 2011, pp. 268–272.

[22] G. Ge, M. Zi, and Q. Junfei, "State-PID feedback control with application to a robot vibration absorber," *International Journal Modelling, Identification and Control*, 2006.

[23] D. Luenberger, "Canonical forms for linear multivariable systems," *IEEE Trans. Autom. Control*, vol. 12, no. 3, pp. 290–293, Jun. 1967.

[24] M. Valášek and N. Olgac, "Efficient eigenvalue assignments for general linear MIMO systems," *Automatica*, vol. 31, no. 11, pp. 1605–1617, 1995.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 142-150 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Into the Online Space: Outcomes for a PASS Online Pilot Across Different First Year University Subjects

Tracey Woolrych[*,1], Melissa Zaccagnini[2], Moira Stephens[3], Murray Stace[4], Rebekkha Middleton[3], Melissa Bergner[2], Sian O'sullivan[5], Reetu Verma[6]

[1]*University of Wollongong; School of Psychology; UOW, 2522, Australia*

[2] *University of Wollongong; PASS; UOW, 2522, Australia*

[3] *University of Wollongong; School of Nursing; UOW, 2522, Australia*

[4] *University of Wollongong; LTC; UOW, 2522, Australia*

[5] *University of Wollongong; Peer Learning; UOW, 2522, Australia*

[6] *University of Wollongong; School of Accounting, Economics & Finance; UOW, 2522, Australia*

ARTICLE INFO

Article history:
Received: 30 September, 2019
Accepted: 24 December, 2019
Online: 22 January, 2020

Keywords:
Peer Assisted Study Sessions
Online Learning
First year University students

ABSTRACT

Peer Assisted Study Sessions (PASS) is a student let program designed to support students transitioning into university and tackling traditionally difficult first year core subjects. The model is collaborative with student leaders facilitating activities and discussions driven by student need. Consistently research has found that students can benefit, in terms of increased grades, from attending PASS, however findings for online delivery modes are mixed. These studies have generally only compared face to face (F2F) with online modes of PASS-like programs for one subject. No study has compared different subjects from varying disciplines to investigate if the benefits of PASS online are the same for all students. PASS at UOW conducted a pilot study of synchronous online sessions, tested across three different first year university subjects. A total of 1.471 students enrolled into these subjects, with 409 attending some form of PASS. Result revealed PASS students gained significantly higher average final marks compared to students who did not attend any type of PASS, regardless off subject. However, results for PASS varied depending upon mode of delivery engaged in (F2F or online) and also with subject. Although not all differences were statistically significant, trends suggest a student/subject interaction that may vary the amount of benefit gained from PASS online formats. Possible drivers for these results are discussed as well as consideration given to cohort effects and student skills for online learning modes.

1. Introduction

As Universities strive to meet the needs of more diverse student populations, online learning options are being rolled out Australia wide. Time poor students, many of whom have families and/or work commitments, require more flexibility to complete their degrees, including a need for online-based supports [1-2]. Peer led learning assistance programs are following this trend by offering synchronous online learning sessions [3]. However it is unclear if students are benefitting from this mode of learning, especially in the first year space where students may not yet be fully independent learners [4]. Currently research into online versions of PASS-like programs have offered positive results in terms of benefits for students. However, many of these studies investigate only one subject at a time, making it unclear whether student and/or subject characteristics may influence the benefits students can obtain.

Research on PASS-like programs has shown many benefits for students including increased grades, particularly in face to face (F2F) delivery modes [5-6]. These findings seem consistent across a range of different courses [7]. However, findings reveal this

[*]Corresponding author: Tracey Woolrych, Email: woolrych@uow.edu.au

consistency is not present for online versions of PASS [8-11]. The F2F environment comes with a range of stimuli and cues which are important in facilitating learning [8]. Given the collaborative nature of Peer Assisted Study Sessions (PASS), these cues become essential for successful student/student and student/leader interactions. There is no guarantee that these can be created or re-created in the online space. Online learning modes may not be able to emulate the F2F student experience [4, 12-14]. One could argue that they should not try, given that the content and mode of delivery for different subjects are intrinsically linked. It is this that may impact on the success of online delivery. Not every subject may be suitable to an online format, with important learning-based elements lost in an online delivery. Therefore it is vital that research address the question of subject and/or student suitability for online delivery, and whether it is beneficial for student learning.

Although current peer-assisted models span a variety of discipline areas, including Engineering and IT [15], languages [8], and occupational therapy [16] few studies have compared these outcomes across multiple subjects and/or disciplines. Subjects vary in their learning outcomes, and the approaches and skill development requirements to meet these. Therefore, a one-size-fits all approach to the adaptation of F2F peer-led learning programs to online versions may not be advisable. Research comparing peer-led online formats across a range of subjects is required to unpack some of the underlying factors that may be contingent on student success. This paper seeks to address this gap by comparing different first year subjects across multiple disciplines. This allows us to explore whether PASS online formats are equally beneficial to students in different subjects and whether other factors such as student and/or subject characteristics may play a role. This is essential to the future development of online formats of PASS-like programs.

2. Peer Assisted Learning (PAL) Programs

Peer Assisted Study Sessions (PASS) was originally based on supplemental instruction (SI) developed by Dr D Martin at the University of Missouri over 30 years ago [17]. SI programs have since spread around the world with many tertiary institutions adopting them in an effort to support their students. PASS have been operating in collaboration with Australian Universities for many years. According to the Supplemental Instruction Journal (2016), SIs are informal sessions whose direction is driven by students and led by students (Leaders) who have previously done well in the subject. Modelled on SI, PASS leaders do not 'teach' as much as facilitate activities and discussions designed to help students more clearly and deeply understand the subject content [17-18]. Students also develop and practice effective study skills [19]. Collaboration is key here, a far cry from the stand and deliver techniques of many lectures: at PASS you will find leaders helping students, and students helping students [20-21].

Research into the peer-led learning space reveals many benefits for students. On average, students who attend these programs obtain higher final marks compared to those that do not attend [5-6]. A higher number of students pass the subject linked with PASS, and more students finally graduate [7]. Dancer et al., [22] also found that there were higher retention rates and few withdrawals in these subjects. Lower-achieving students gained more benefits in terms of grade increases compared to higher-achieving students [22]. The positive benefits that students can gain from PASS-like programs are clear. What is less clear is how these peer-led models of learning translate into the online space.

2.1. Findings in the online space

Increasingly today's students are facing time pressures, trying to balance work/family and study commitments. This results in students desiring more flexible study options at University. Education providers are meeting this need through the delivery of blended and online learning [23]. Online study is offered over a wide range of subjects and courses. Peer led assisted programs are following suit, offering online, mostly synchronous, study sessions for students.

Research into the benefits and success of peer lead online learning is still growing, and the literature is not of a single voice in its success. Some studies have found student benefits in the form of increased satisfaction and higher average grades e.g.:[24-28]. Students also report reduced feelings of isolation [26], enjoying the flexibility [20] and collaboration [5, 29-30] that resulted from the online sessions. Online formats of peer lead learning resulted in increased student interest and engagement [30], as well as increased understanding [29]. In particular, those subjects that required skills development in IT [15] and languages [8, 31] reported success in the online mode.

Some studies revealed no differences in average final marks between students attending F2F peer-led learning sessions, compared to online attendance [8, 26-28] suggesting, on the surface, that both modes afford students the same benefits. However other studies have found a lack of grade increases for their online cohorts [6]. These variations in results could be explained by poor student uptake for the online formats [13, 23, 32] and a lack of student participation in feedback [33]. Technical issues experienced in the online format may also result in less that optimal learning environments resulting in some students abandoning the sessions altogether [32]. Recent studies have also highlighted the need for leader training specifically for the online mode of delivery [13, 23, 32]. Terminology is also an issue in this area of research. The definitions of PASS attendance are variable which may explain some of these results. As Dawson et al., [7] point out, the distinction between peer-groups and non-peer groups is at times arbitrary with studies using cut-offs varying from 1, 3 or more classes to define group conditions. The variation and rapidly evolving nature of the technology being used may also contribute to variable findings, with online studies needing to be read in the broader context of the technology "point in time".

The lack of gains achieved in online learning formats may also be the result of insufficient student skills. These studies often involve first year university students, some of whom struggle with the self-reliant and self-driven nature of online formats [4]. They may not yet have development the necessary problem solving [34], time-management and self-management skills [35, also see 3], such as motivation and commitment, necessary to succeed [36-38]. The online learning format presents a large departure from traditional learning environments many students would have been exposed to in their prior learning. Mature age students, on the other hand, may lack the necessary digital literacy and confidence to successfully navigate the online learning environment [39-42].

Studies on blended learning models report push-back from some students against taking responsibility for their own learning [14, 43]. Students new to university may have had little experience in online learning and so lack the relevant skills required for success.

An alternative explanation may be that the different subjects attract different types of students. Therefore, there may be unseen cohort effects driving some outcomes. Few studies have compared multiple subjects from different disciplines.

Another possible explanation may be that not all subjects are suited to an online mode of delivery. Whilst peer-led learning has been conducted across a range of subjects and disciplines, it should be remembered that subjects vary in their learning outcomes, approaches and requirements, meaning a one-size fits all approach may not be appropriate. The literature is unclear on this, as some studies do not state the subject under investigation, nor the faculty within which the subject is situated. The findings do suggest that STEM (science, technology, engineering, and maths) subjects do well in the online format [7, 44], although some studies did not find any grade increases for those cohorts [30, 33]. PASS conducted for Biology, as both a STEM and nursing subject, resulted in higher grades for students who attended in both the F2F and online modes [27-28]. Other subjects that have had successful outcomes in online formats are pre-doctoral (dentist) studies [24], and Education [29]. Dennis' [45] study found no differences in average grades between F2F and online versions of PASS for physical therapy students. Overall the literature suggests that students of most subjects can benefit from PASS-like programs, however whether these benefits successfully translate into the online space is less clear. In order to identify the factors that influence PASS-online success, studies need to investigate differences or similarities in outcomes across different subjects, across various disciplines.

3. Peer Assisted Study Sessions (PASS) at University of Wollongong (UOW)

The University of Wollongong (UOW) houses the Australian National Centre for PASS. The PASS program was launched in 2002 and in the intervening years has met with much success [46]. Students report higher grades and increased confidence. Traditionally, PASS is run in parallel with a range of first year subjects to support students with their transition to University study. Attached commonly to challenging first year core subjects, PASS is non-remedial, with PASS leaders guided by student needs and concerns. Currently PASS at UOW facilitates over 40 subjects (some in multiple semesters) across all faculties and onshore campuses.

The study sessions are run by PASS leaders, who have successfully passed the subject themselves. The sessions are driven by the attending students who choose the subject content to be focused on for discussion, clarification, or practice. PASS leaders engage students with a range of learning activities and guide students to develop study strategies. Activities in the F2F sessions can include brainstorming, concept lists, filing in the missing elements of an equation, and diagram labelling. The learning environment is collaborative between students and leaders. Students are encouraged to share their thoughts and knowledge with others, sharing answers and processes, often teaching back to other students. PASS students, in this way, have a direct stake in the teaching process, enhancing their own learning experiences.

3.1. The current project – PASS Online Pilot.

At UOW the demand for PASS is high. Large first year student cohorts, as well as time and space availability, limit the number of sessions that can be offered for each subject. Students also voiced a desire for some out-of-hours sessions to accommodate work and/or family commitments. Regional students also needed a way to access PASS in the absences of a suitable leader being available on their campus. In response to this, PASS at UOW trialled an online format using Blackboard Collaborate, an online synchronous platform.

PASS is student led and student focused. The selection of an online platform required careful consideration of this. The aim of the pilot was to emulate, as much as possible, the F2F student experience. It was not the aim of the pilot to replace the usual F2F sessions, but rather supplement them, by giving students more flexibility of access. The online environment needed to align with the core values and practices of PASS.

In the spirit of the student-driven principals that lie at the heart of the PASS model, student leaders were recruited to test different synchronous online platforms, as well as host the online sessions during the pilot. Blackboard collaborate was chosen for its flexibility, ease of use, and ease of access (embedded in Moodle). The platform provides for online synchronous discussion spaces, student/student and leader/student interactions, and exchange of document-based information including power point slides.

PASS leaders tested the capabilities of the system through role-play, swapping between student and leader roles to gain a deep understanding of the student experience. These experiences were shared with other leaders as part of their training for the online system and formed the basis for written guides for use in future training.

Three large cohort subjects from three different faculties were chosen for the pilot which first ran in second session, 2017. All were first year core subjects and traditionally considered difficult and challenging by students. These included: COMM121 (Statistics for Business, Commerce); NMIH106 (Essentials of Care A, Nursing); and PSYC123 (Research Methods and Statistics, Psychology).

It was predicted that if the online version of PASS provided similar benefits to F2F sessions, then, regardless of mode, PASS students would all obtain higher average marks compared to students who did not attend PASS. It was hypothesised that there would be no statistically significant difference in average mean mark achieved between the different PASS delivery modes. Finally, it was hypothesised that, regardless of PASS mode, the more hours dedicated to PASS the greater the benefits for students in the form of higher final marks.

Subject Overviews

Two of the piloted subjects were statistics-based. COMM121 and PSYC123 are first year introductory, core, statistics subjects in the schools of Business and Psychology respectively. COMM121 deals with quantitative statistical analysis for data in

the business and economics environments. Analyses and concepts include probability; hypothesis testing; correlations; regression; and time series forecasting. Students also develop skills in the use of excel and PHStat2 programs. COMM121 is an applied subject, teaching students how to use statistics to answer questions about business success and business-based weaknesses.

PSYC123 is also a quantitative-based statistics subject designed to help students answer questions about the human condition. A core subject for Psychology and Social Science students in the faculty of Social Science, students learn concepts relating to theoretical distributions; sampling; descriptive and inferential statistical analysis (z-scores; correlation; t-tests; chi-square); hypothesis testing; and methodological considerations, including validity and reliability. Students conduct all analysis and calculations by hand with a strong emphasis on result interpretation.

Essentials of Care A (NMIH106) focuses on the skills essential for successful patient/client interactions, including communication; person-centred care; knowledge (intervention; evidence; and health science); skills practice; and documentation. The subject builds directly upon knowledge and skills gained in first session nursing subjects. Foundational science knowledge is at the core of this subject, how it relates to nursing practice and how it informs patient/client communications with other health care personnel.

4. Methods

4.1. Participants

A total of 1,471 students were enrolled across the three subjects. Students were deemed PASS participants if they attended at least on session in any mode. Those students who did not obtain a final mark or obtained a final mark of zero (0) were excluded from analysis, leaving a sample size of 1,423. Of these, 409 students attended a form of PASS across the three subjects (see table .1. below).

4.2. Procedure

PASS in the online format was run concurrently with F2F sessions and with normal subject delivery in each school, throughout the second teaching session of 2017, and was open to all enrolled students from the three subjects piloted. Students voluntarily signed up by enrolling in F2F or online PASS, or a combination of both. Students were briefed about the online pilot which outlined some of the potential benefits of the program and expected time commitments (for feedback via online survey). Students could attend as many or as few PASS as they wished, in any mode of delivery. During week 6 of session, and again in week 13, students were invited to complete a brief Survey-Monkey questionnaire to gauge their experiences with the online format. All data from the questionnaires, and end of session marks for all students in each subject, were de-identified to protect anonymity. This project had ethics approval from UOW (2017/263).

4.3. Analysis

Participants were split into four discrete groups, dependent upon the PASS mode engaged in. These were PASS F2F only (PF2F); PASS online only (PO); PASS F2F and online in combination (PO+F2F); and no PASS attendance (NP). Sample numbers in the PASS online formats were very low. Non-parametric tests yielded the same results as parametrics, therefore it is the parametric test results given here. Significant results, then, should be interpreted with caution.

5. Results

5.1. Descriptives

A total of 409 students attended PASS across all three subjects. Students in PSYC123 (m =69.71) achieved the highest grades overall, followed by NMIH106 (m =66.53) then COMM121 (m =59.12). These differences were all statistically significant ($F(2,1420)=44.69$, $p<.001$ Students who attended PASS, in any format, achieved higher average marks compared to those that did not attend. For students who attended PASS, PSYC123 students achieved the highest grades (m=73.19; CI[71.10;75.28]), followed by COMM121 students (m=70.35, CI[67.00;73.69]) and NMIH106 (m=69.39, CI[66.96;71.82]) (see table .1. below).

For COMM121, 20% of students attended only 1 session of PASS, compared to 34% of students in PSYC123, and 40% of students in NMIH106.

5.2. PASS modes

Students who attended any mode of PASS achieved higher average final marks (m = 71.24) compared to students that did not attend (m = 62.35; $t(970.462)=-9.069$, p <.001; CI [-10.82; -6.97]; see figure 1 below).

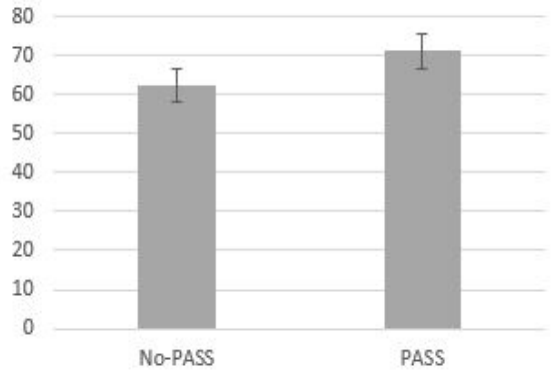


Figure 1: Mean final mark comparisons between No-PASS and PASS groups (regardless of PASS delivery mode).

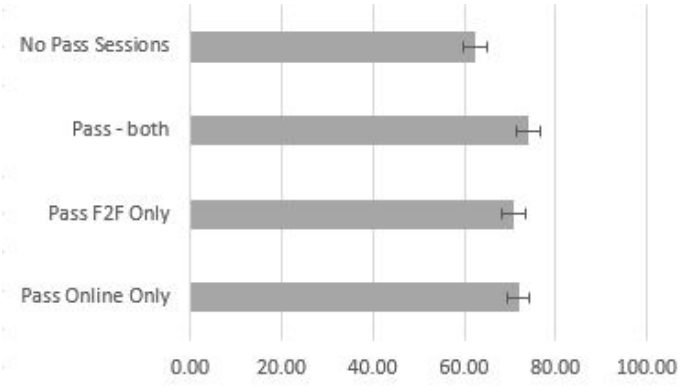


Figure 2: Mean final mark distributions comparing No-PASS and all modes of PASS delivery

Table 1: Mean final grade and mean hours of PASS attended across all PASS modes, No-Pass and for each subject.

Subject	PASS mode	n	Mean Final Grade	SD	CI	Mean hours	SD	CI
COMM121	No-PASS	411	55.81	23.13	[53.57;58.05]	-	-	
	PO	9	67.67	7.89	[61.60;73.73]	4.22	5.07	[.33;8.12]
	F2F	97	69.78	19.59	[65.83;73.73]	5.36	3.63	[4.63;6.09]
	PO+F2F	15	75.60	16.09	[66.69;84.51]	9.40	4.94	[6.66;12.14]
	Overall	***532***	*59.12*	*22.98*	*[57.16;61.07]*	*5.78*	*4.13*	[5.07;6.48]
NMIH106	No-PASS	258	65.21	12.55	[63.67;66.75]	-	-	
	PO	9	74.22	8.15	[67.96;80.49]	1.00	-	[1.00;1.00]
	F2F	104	68.72	13.63	[66.07;71.37]	3.60	2.68	[3.07;4.12]
	PO+F2F	6	73.67	14.75	[58.19;89.14]	5.33	2.16	[3.07;7.60]
	Overall	***377***	*66.53*	*12.95*	*[65.22;67.84]*	*3.49*	*2.67*	[3.00;3.97]
PSYC123	No-PASS	345	68.00	17.80	[66.12;69.89]	-	-	
	PO	15	73.00	20.23	[61.79;84.20]	4.13	3.78	[2.04;6.22]
	F2F	137	73.26	12.83	[71.09;75.42]	5.86	4.04	[5.18;6.54]
	PO+F2F	17	72.82	15.04	[65.09;80.56]	9.88	5.40	[7.11;12.66]
	Overall	***514***	*69.7082*	*16.74*	*[68.26;71.16]*	*5*	*4.36*	[5.48;6.74]
		1423						

The highest grades were achieved when students attended a combination of PASS online and F2F sessions (see figure 2 below). However, there was no statistically significant difference in grades between the different PASS delivery modes ($F(2,408)=.772$, $p=.463$) suggesting that all students benefitted from PASS, regardless of mode.

There were variations between subjects in average final marks for students who attended PASS ($F(2,408)=2.47$, $p=.086$), with PSYC123 PASS students obtaining the highest marks (see table 1 above). This difference is close to statistical significance, however, low sample sizes in some of the different PASS modes may underpower this result.

5.3. PASS modes by subject

For COMM121, PSYC123 and NMIH106 there were no statistically significant differences in average student marks across the different PASS modes. Trends indicated that, dependent upon the subject taken, the amount of mark increase varied dependent upon the mode of PASS undertaken (see figure 3 below).

For COMM121 students, the highest marks were obtained by students attending a combination of PO and F2F sessions ($m=75.60$), followed by F2F only ($m=69.78$) and then Online only ($m=67.67$). Conversely, students of NMIH106 gained higher marks when attending Online only versions of PASS ($m=74.22$) or a combination of PO and F2F ($m=73.67$), the lowest marks were achieved by students who attended F2F sessions only ($m=68.72$). PSYC123 students performed almost equally across all three delivery modes of PASS: PO (m=73.00); F2F (m=73.26); and PO+F2F (m=72.82). Whilst these differences were not statistically significant, it is interesting to note that students in different subjects gained varied benefits from the different PASS delivery modes.

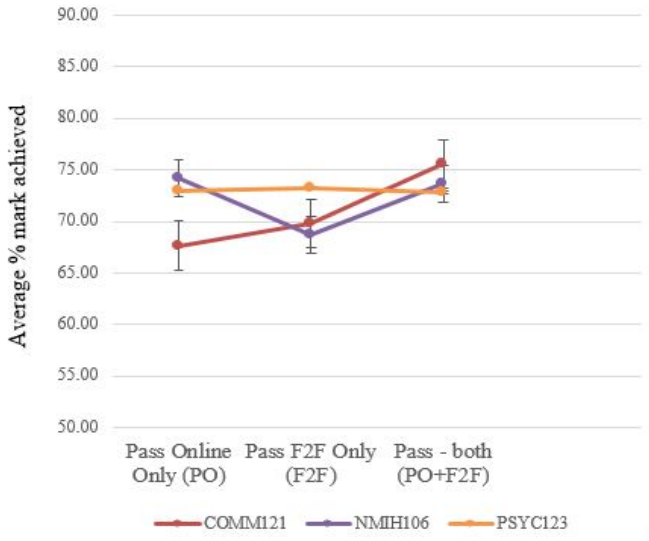


Figure 3: Comparisons of final average mark by subject and PASS delivery mode.

5.4. Correlations

To investigate the impact of hours of PASS, in its various modes, on final grade outcomes a series of correlations were conducted. Not surprisingly, there was a moderate positive, statistically significant result for number of PASS hours completed and final average marks via Pearson's Correlation Coefficient, ($r(=.226, n=409, p < .001)$. Therefore, regardless of PASS delivery

mode, the more hours completed by students, the higher their final mark.

This positive relationship held true for both NMIH106 (r=.25, n=-119, p=.006) and PSYC123 (r=.25, n = 169, p = .001). For COMM121 however the relationship was not significant (r = .16, n=121, p=.079). Looking into these correlations further, only positive relationships between hours and final marks were found for the F2F mode of PASS for NMIH106 (r=.28, n=104, p.004) and PSYC123 (r= .38, n=137, p < .001). All online modes returned non-significant results suggesting that for online versions students did not benefit from longer hours in PASS. The lack of student numbers in some of the online cohorts in this study, however, may have contributed to the non-significant results.

There were some negative relationships found, although they were non-significant. For COMM121 and PSYC123 students, hours spent in the PO+F2F mode of PASS was negatively correlated with final mark achieved, meaning that more hours did not necessarily equate to higher marks.

6. Discussion

Increasingly, institutions are offering online versions of programs to meet student demand for flexibility. Peer-led learning models have been found to benefit students in terms of final marks [5-6,22] as well as confidence [3]. However, whether these benefits can be successfully transferred to online versions is less clear. As Wang et al., [12] point out, online formats may not be able to replicate the stimuli and cues that are essential to interpersonal interactions and collaboration that can be found in F2F versions. Inconsistent results in the literature suggest this is a complex question [8] meaning other factors might be of influence. The findings from this Online PASS pilot highlight the fact that, depending upon the mode of PASS engaged in, student benefits in terms of final grades can vary. Students characteristics, which may help explain this variation, were not investigated in this current study, and should be the focus of future research.

Overall, students who attended PASS achieved significantly higher final marks compared to students who did not attend. This result seems to be driven by the F2F version of PASS, although there were no significant differences in final marks between PASS delivery modes. This is in line with previous literature that has highlighted the benefits of Peer-led learning programs [3, 24-26]. Therefore, PASS can result in positive benefits for students across a range of subjects, for the F2F version at least.

The results varied according to subject. For NMIH106 and PSYC123 there were benefits for students in the online versions of PASS, however COMM121 students seemed to struggle with this. In COMM121 in particular, numbers in the online cohorts were low which may have resulted in underpowering of the statistics used. This is not unusual according to the literature, where, despite student demand for more flexible learning options, when offered, the uptake is often low [20,32,46].

This study also found that overwhelmingly students still prefer the F2F versions of PASS. Students straight from high school, and mature age students, may be less adept at learning in the online space, preferring the more familiar class-room-like environment. This may be due to a lack of confidence, knowledge and/or skills for online formats [3-4, 34-38]. Skills development is an important issue that should be addressed for students coming into University study. Enhancing students' online learning skills will help to build their confidence in the online space, making them better able to take advantage of PASS online.

The nursing subject, NMIH106, was the only subject where students in the online only condition achieved higher marks than students in the F2F only condition. Again, this highlights the fact that different subjects are host to students of varying characteristics that can impact the level of success they achieve in online learning formats. This finding is counter to much of the literature which indicate either F2F modes result in higher grades, or that there is no difference with the online version. Although the current results were not statistically significant, it was the only trend where average final marks were higher for the online versions compared to F2F. This may be reflective of a cohort effect. Nursing students in Australia tend to be older than most other first year university students [48] with a mean age of 28.6. Salamonson et al., [48] also reported that this cohort work an average of 13.6 hours a week, although it is unclear if this is above or below the average for other first year university students. Although age and work hours were not elicited in this study it may be reasonable to suggest that the UOW Nursing cohort are not unlike that of the Salamonson et al.'s study [48]. The fact that this group performed better in the online versions of PASS perhaps speaks to more mature time-management and self-motivation skills. The availability of an online peer assisted learning option may also be particularly helpful in overcoming barriers to scheduling F2F PASS in conjunction with compulsory practical placements for this cohort and their peer leaders. These factors should be investigated further.

Interestingly, the outcomes for both COMM121 and PSYC123 were quite different, despite both being first year statistics subjects. This suggests that different student cohorts may bring with them a different set of characteristics which may influence the success of online learning. Some students may be more prepared and have better time-management and self-motivation skills that are essential for this type of learning [3-4, 34-38]. More studies comparing subject cohorts across years should be conducted in order to clarify if the effect is due to characteristics as a result of subject/degree selection or a yearly cohort effect.

The hours spent at PASS was, in most cases, positively related to final marks, which is consistent with previous research [44]. However, for both COMM121 and PSYC123 students in the PO+F2F condition a negative relationship between hours and final marks was found, although it was non-significant. Low participant numbers in the online conditions again may have contributed to the non-significant result, however it is an interesting trend. The negative relationship is difficult to explain except to say that other factors may be involved that were not the subject of this study. Both COMM121 and PSYC123 are statistics subjects, meaning students who attended both F2F and online PASS may have lacked confidence with the subject. It may also be the case that these students lacked the necessary skills to succeed, despite their PASS attendance. The role of PASS alongside other learning supports should be investigated further.

Self-selection bias is a common problem in peer-led learning program research, as it is often the most diligent students that

attend [see 7, 49-50]. Students may participate in PASS for a number of reasons, including the drive to gain higher marks, or lack of confidence in the subject. Fetner [4] made the observation that some students struggle with the self-driven nature that is often inherent in the online learning space. Despite the flexibility this mode of learning affords students, a lack of time-management and self-motivation means those students may not be maximising the benefits that are on offer. This is particularly relevant for first year university cohorts who have not yet found the most efficient and effective study methods [43]. It has also been reported that students may push-back against online learning models, believing that the responsibility for their learning lies with the educators, rather than with themselves [14]. Therefore, it is imperative that students be offered the opportunity to develop the skills that will help them succeed in the online learning space.

Overall, the current study indicates that F2F versions of PASS deliver consistent benefits to students via increased grades, across a variety of subjects. The benefits to be had in the online learning modes, however, are less clear. The current study highlights the need to further investigate student and/or subject characteristics that may impact the benefits achieved by students engaging with the online format. Other factors such as student skills, maturity, and confidence, may be important factors that impact online learning outcomes. Similarly, subject characteristics should be studied which may reveal factors regarding the suitability of different subjects for this particular learning platform.

6.1. Limitations

Low student numbers in many of the online conditions may have undermined some of the inferential statistical analyses undertaken, therefore results should be interpreted with caution. Low sample sizes in online conditions is a problem in the literature [20,32,46], where student uptake has been lacking. This problem can be overcome in the future through higher sample numbers, however, given that enrolment into online conditions is voluntary it is likely that this will continue to be a problem for future research. There are positive outcomes from this study that suggest, with the right supports and skills development, students can gain equal benefit from online versions of PASS as those obtained in F2F versions.

The high number of students who attended only 1 PASS may also have clouded results, underpowering the analysis, particularly for the correlations. Whilst some studies do not include students that have attended only 1 PASS class [7], it was decided that any student that attended any mode of PASS for at least one session be included in order to reduce the impact of low numbers in some of the online conditions.

It should also be noted that this was a Pilot study. Students at UOW had not previously had the opportunity to participate in PASS in the online format. Despite receiving specific training for the online space, PASS leaders themselves were learning along the way. Sessions at the beginning of the academic period may have been subject to technical issues, particularly in the statistics subjects where difficulty in sharing formulas and the creation of specific mathematical symbols was problematic. Therefore, the focus, in these early sessions, may have veered away from PASS activities and interactions to the implementation issues and student user troubleshooting. PASS leaders reported anecdotally that they, and their students, grew in confidence in working with the online system as the semester progressed. These lessons have now been included in the training guides and manuals for future PASS online leaders.

7. Conclusion

This study may explain some of the inconsistencies in findings for the Peer-led online learning space. Students from different subjects had differing outcomes dependent upon the type of PASS undertaken. This suggests that there may be underlying student characteristics which vary according to the subject taken, which may impact successful learning in the online space. Larger sample sizes, particularly for the online conditions, will help to clarify this, as will studies comparing results for the same subjects across years. Student characteristics such as work and/or family commitments, age, and other factors should also be investigated.

Students may also lack the necessary skills to succeed, particularly in the dimensions of self-motivation and time-management. Despite students reporting the online system easy to use, most students still showed a preference for the F2F format. This highlights the need for student skills development in the first-year space so that students can more confidently and effectively take advantage of the flexible learning that PASS online can offer.

PASS can help students achieve higher marks. Overwhelmingly students that attended PASS F2F sessions achieved significantly higher marks compared to those who did not attend. There were a high number of students that attended only 1 session of PASS and it is unclear why this might be the case. What is clear is that dedication to PASS, in the form of multiple sessions, has benefits for students.

The study provides a useful contribution to the understanding of PASS style learning environments in the online mode. Consideration of subject selection, student demographics, technology available, training of peer leaders, and digital literacy of participants are areas for further investigation. Future studies should also investigate student and subject characteristics that may impact the benefits students gain from PASS. This study, limitations considered, indicates that not all subjects may be suitable for an online PASS platform. The role of peer assisted learning in online synchronous modes and how this relates to asynchronous interactions is outside the scope of this study, however, is an area for further investigation.

Peer-led learning programs are essential in helping students transition into university study. This PASS pilot demonstrates that, for traditionally difficult first year subjects, PASS can give students the winning edge they need. In this student-driven environment, students learn not only from PASS leaders, but from other students, engaging in discussions and activities that help elucidate the more difficult concepts in a subject, and also develop the study skills necessary to succeed. Venturing into the online space for Peer-led programs comes with it challenges, however with future studies investigating student characteristics and other factors that impact online learning outcomes will help ensure the future success of PASS Online.

8. References

[1] Gillies, D. “Student perspectives on videoconferencing in teacher education at a distance” Distance Education, 29(1), 107–118, 2008.

[2] Kearsley, G. "Is online learning for everybody?" Educational Technology, 42(1), 41-44, 2002. Retrieved from http://www.jstor.org/stable/
[3] McKenzie, K., & Schweitzer, R. "Who succeeds at university? Factors predicting academic performance in first year Australian university students" Higher Education Research & Development, 20(1), 21-33, 2001.
[4] Fetner, M. "What do unsuccessful online students want us to know?" Journal of Asynchronous Learning Networks, 17(1), 13-27, 2013.
[5] Pereira, R. "UniPASS Report Semester 1 2012," Curtin University: Australia,2012. https://life.curtin.edu.au/local/docs/2012_S1_UniPASS_Report_Web_Edit.pdf
[6] Taylor, K., & Kelly, T. "Comparison of effectiveness of several peer learning programs to support 1st year chemistry" in the 2014 Australian Conference on Science and Mathematics Education, 89-90, September 2014. (formerly UniServe Science Conference).
[7] Dawson, P., van der Meer, J., Skalicky, J. & Cowley, K. "On the effectiveness of Supplemental Instruction: A systematic review of Supplemental Instruction and Peer-Assisted Study Session Literature between 2001 and 2010" Review of Educational Research, 84(4), 609-639, 2014.
[8] Yu-Chen, K., Hui-Chun, C., & Chi-Hao, H. "A learning style-based grouping collaborative learning approach to improve EFL students' performance in English courses" Journal of Educational Technology & Society, 18(2), 284-298, 2015.
[9] Brandon, D., & Hollingshead, A. "Collaborative learning and computer-supported groups" Communication Education, 48(2), 109-126, 1999.
[10] Spaniol- Mathews, P., Letourneau, L.E., & Rice, E. "The impact of online supplemental instruction on academic performance and persistence in Undergraduate STEM courses" Supplemental Instruction Journal, 2(1), 19-32, 2016.
[11] Smith, J., Wilson, S. B., Banks, J., Zhu, L., & Varma-Nelson, P. "Replicating Peer-Led Team Learning in cyberspace: Research, opportunities, and challenges" Journal of Research in Science Teaching, 51(6), 714-740, 2014.
[12] Wang, Q., Huang, C., & Quek, C.L. "Students' perspectives on the design and implementation of a blended synchronous learning environment" Australian Journal of Educational Technology, 34(1), 1-13, 2018.
[13] Beaumont, T.J., Mannion, A.P., & Shen, B.O. "From the campus to the cloud: The online Peer Assisted Learning Scheme" Journal of Peer Learning, 5(1), 1-15, 2012. http://ro.uow.edu.au/ajpl/vol5/iss1/6
[14] O'Flaherty, J., & Phillips, C. "The use of flipped classrooms in higher education: A scoping review" Internet and Higher Education, 25, 85-95, 2015.
[15] Huijser, H., & Kimmins, L. "PALS online and community building: A contradiction in terms?" in the ASCILITE 2005: 22nd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Balance, Fidelity, Mobility-Maintaining the Momentum? (1). Queensland University of Technology, Teaching and Learning Support Services, 2005.
[16] Larkin, H,., & Hitch, D. "Peer Assisted Study Sessions (PASS) preparing occupational therapy undergraduates for practice education: A novel application of a proven educational intervention" Australian Occupational Therapy Journal, 55, 100-109, 2019. doi: 10.1111/1440-1630.12537
[17] Topping, K. J. "The effectiveness of peer tutoring in further and higher education: A typology and review of the literature," Higher Education, 32(3), 321-345, 1996.
[18] Longfellow, E., May, S., Burke, L., & Marks-Maran, D. "'They had a way of helping that actually helped': a case study of a peer-assisted learning scheme" Teaching in Higher Education, 13(1), 93-105, 2008. doi: 10.1080/13562510701794118
[19] Capstick, S., & Fleming, H. "Peer Assisted Learning in an Undergraduate Hospitality Course: Second Year Students Supporting First Year Students in Group Learning." Journal of Hospitality, Leisure, Sport and Tourism Education 1(1), 69–75, 2002.
[20] Brown, K., Nairn, K., van der Meer, J., & Scott, C. "'We were told we're not teachers….it gets difficult to draw the line': Negotiating roles in Peer-Assisted Study Sessions (PASS)" Mentoring & Tutoring: Partnership in Learning, 22(2), 146-161, doi: 10.1080/13611267.2014.902559
[21] Cusick, A., Camer, D., Stamenkovic, A., & Zaccagnini, M. "Peer Assisted Study Sessions for research trainees" Journal of Peer Learning, 8(14),18-33, 2015. https://ro.uow.edu.au/ajpl/vol8/iss1/4
[22] Dancer, D., Morrison, K., & Tarr, G. "Measuring the effects of peer learning on students' academic achievement in first-year business statistics" Studies in Higher Education, 40(10), 1808-1828, 2015.
[23] Lim, C., Anderson, F., & Mortimer, R. "PASSport to the Cloud- Results of a Peer-Assisted Study Sessions (PASS) Online Pilot Program" Student Success, 7(2), 59-64, 2016.
[24] Carter-Hanson, C., & Gadbury-Amyot, C. "Implementing Supplemental Instruction online to create success in high-stakes coursework for pre-doctoral dental students" Supplemental Instruction Journal, 2(1), 53-75, 2016.
[25] Court, S., & Molesworth, M. "Course-specific learning in peer assisted learning schemes: A case study of creative media production courses" Research in Post-compulsory Education, 13, 123-134, 2008.
[26] Evans, M.J., & Moore, J.S. "Peet tutoring with the aid of the internet" British Jouranl of Educational Technology, 44(1), 144-155, 2013. doi:10.1111/j.1467-8535.2011.01280.x
[27] Finlay, K., & Mitchell, M. Z. "Evaluating the effectiveness of online vs. face-to-face Supplemental Instruction in introductory biology courses" unpublished, University of Regina, 2017.
[28] Hizer, S. E., Schultz, P., W & Bray, R. "Supplemental Instruction Online: As Effective as the Traditional Face-to-face model?" Journal of Science, Education & Technology, 26(1), 100-115, 2017.
[29] Bone, J., & Edwards, S. "Connecting in rhizomic spaces: Peer-assisted learning (PAL) and e-learning in teacher education" Australian Journal of Adult Learning, 55(1), 54-74, 2015.
[30] Edwards, S., & Bone, J. "Integrating peer assisted learning and eLearning: Using innovative pedagogies to support learning and teaching in higher education settings" Australian Journal of Teacher Education, 37(5), 1-12, 2012.
[31] Dekhinet, R., Topping, K., Duran, D., & Blanch, S. "Let me learn with my peers online! Foreign language learning through reciprocal peer teaching" Innovate: Journal of Online Education, 4(3), 2, 2008.
[32] Watts, H., Malliris, M., & Billingham, O. "Online Peer Assisted Learning: Reporting on practice" Journal of Peer Learning, 8(8), 85-104, 2015.
[33] Rourke, L., & Anderson, T. "Using peer teams to lead online discussions" Journal of Interactive Media in Education, 2002, ART.1, 2002.
[34] McPherson, M., & Nunes, M.B. "The failure of a virtual social space (VSS) designed to create a learning commuinity: Lessons learned" British Journal of Educational Technology, 35, 305-321, 2004.
[35] Lew, M.D.N., Alwis, W.A.M., & Schmidt, H.G. "Accuracy of students' self-assessment and their beliefs about its utility" Assessment & Evaluation in Higher Education, 35(2), 135-156, 2010. doi: 10.1080/02602930802687737
[36] Brouwer, J., Jansen, E., Hofman, A., & Flache, A. "Early tracking or finally leaving? Determinants of early study success in first-year university students" Research in Post-Compulsory Education, 21(4), 376-393, 2016. doi: 10.1080/13596748.2016.1226584
[37] MacCaskill, A., & Denovan, A. "Developing autonomous learning in first year university students using perspectives from positive psychology" Studies in Higher Education, 38(1), 124-142, 2013. doi:10.1080/03075079.2011.566325
[38] Mason, H.D. "Grit and academic perofrmance among first-year university students: A brief report" Journal of Psychology in Africa, 28(1), 66-68, 2018. doi:10.1080/14330237.2017.1409478
[39] Hryciw, D.H., Tangalakis, K., Supple, B., & Best, G. "Evaluation of a peer mentoring program for a mature cohort of first-year undergaduate paramedic students" Adv Physiol Educ, 37, 80-84, 2013, doi:10.1152/advan.00129.2012
[40] Cullity, M. "Challenges in understanding and assisting mature-age students who participate in alternative entry programs" Australian Journal of Adult Learning, 46(2), 175-201, 2006
[41] Spedding, J., Hawkes, A.J., & Burgess, M. "Peer Assisted Study Sessions and student performance: The role of academic engagement, student identity, and statistics self-efficacy" Psychology Learning & Teaching, 16(1), 144-163, 2017. doi: 10.1177/1475725716687166
[42] Seiber, V. "Diagnostic online assessment of basic IT skills in 1st-year undergraduates in the Medical Sciences Division, University of Oxford" British Journal of Educational Technology, 40(2), 215-226, 2009. doi:10.1111/j.1467-8535.2008.00926.x
[43] Smart, K.L., & Cappel, J.J. "Students' perceptions of online learning: A comparative study" Journal of Information Technology Education, 5, 201-219, 2006.
[44] Ng, R., Kaur, A., Sheikh Mohamed, S. F., Latif, L. A., & Bahroom, R. "E-mathematics: pre-instructional and supplement instruction and their impact of student's online participation and final exam score" Asian Association of Open Universities Journal, 4(1), 27-36, 2009.
[45] Dennis, J. K. "Problem-based learning in online vs. face-to-face environments" Education for Health, 16(2), 198-209, 2003.
[46] Zaccagnini, M., & Verenikina, I. "Peer assisted study sessions for postgraduate international students in Australia" Journal of Peer Learning, 6, 86-102, 2013.http://ro.uow.edu.au/ajpl/vol6/iss1/8
[47] Nikolic, S. & Nicholls, B. "Exploring Student Interest of Online Peer Assisted Learning Using Mixed-Reality Technology" Advances in Intelligent Systems and Computing, 715, 48-54, 2017.

[48] Salamonson, Y., Ramjan, L., Lombrado, L., Lanswer, L.D., Fernandez, R., & Griffiths, R. "Diversity and demographic heterogeneity of Australian nursing students: a closer look" International Nursing Review, 59(1), 59-65, 2011.
[49] Kochenour, E.OD.S., Jolley, J.G., Kaup, D.L., Roach, P.K.D., & Wenzler, L.A. "Supplemental Instruction: An effective component of student affairs programming" Journal of College Student Development, 38(6), 577-586, 1997.
[50] MCarthy, J. "Blended learning environments: Using social networking sites to enhance the first year epxerience" Australiasian Journal of Educational Technology, 26(6), 729-740, 2010.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 151-158 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

SiC-FET Gas Sensor for Detecting Sub-ppm Gas Concentrations

Yoshitaka Sasago[*, 1], Hitoshi Nakamura[1], Takahiro Odaka[1], Atsushi Isobe[1], Shigenobu Komatsu[1], Yohei Nakamura[1], Taizo Yamawaki[1], Chiko Yorita[2], Nobuyuki Ushifusa[2], Kohei Yoshikawa[3], Kazuo Ono[1], Yumiko Anzai[4], Shuntaro Machida[1], Masaharu Kinoshita[1], Koji Fujisaki[1], Kenji Okishiro[5], Yuta Sugiyama[6]

[1]*Center for Technology Innovation – Electronics, Hitachi, Ltd., Tokyo, 185-8601, Japan*

[2]*Center for Technology Innovation – Production Engineering, Hitachi, Ltd., Kanagawa, 244-0817, Japan*

[3]*Center for Technology Innovation – Materials, Hitachi, Ltd., Ibaraki, 319-1292, Japan*

[4]*Center for Technology Innovation – Healthcare, Hitachi, Ltd., Tokyo, 185-8601, Japan*

[5]*Global Research & Innovative Technology Center GRIT, Hitachi Metals, Ltd., Saitama, 360-8577, Japan*

[6]*Cable Materials Company, Hitachi Metals, Ltd., Ibaraki 319–1418, Japan*

ARTICLE INFO

Article history:
Received: 30 September, 2019
Accepted: 23 December, 2019
Online: 22 January, 2020

Keywords:
Capacitor
Platinum
YSZ
Nickel oxide
Nitrogen oxides (NOx)
Nitrogen monoxide (NO)
Oxygen (O_2)
Sulfur oxide (SO_2)
Carbon monoxide (CO)
Ammonia (NH_3)

ABSTRACT

This paper describes a silicon carbide-field effect transistor (SiC-FET) gas sensor that is able to detect NO, O_2, NH_3, CO, and SO_2. The gate of the sensor FET is a gas detection layer that consists of yttria-stabilized zirconia, nickel oxide, and platinum. The threshold voltages of the sensor depend on the target gas concentration, measurement temperature, and sensor gate materials. The device on SiC substrate can measure gas at temperatures up to 800°C. The gate material dependence and the wide-range temperature dependence of the sensitivity make it possible to detect the concentration of each component in a mixture of gases. In addition, a gate stack composed of composite materials improves the sensitivity of the FET sensor by helping to maintain the sensor's heat resistance and the reproducibility of its measurements.

1. Introduction

In particular, pollution in the atmosphere directly affects human health. Here, nitrogen oxides (NOx) are the most harmful components of automobile exhaust gas, and sensitive gas sensing is required for detecting them. In Ref. [1], we developed an SiC-FET-type NOx sensor for high-temperature exhaust gas with a sub-ppm nitrogen monoxide (NO) detection capability.

Limiting-current-type [2], mixed-potential-type [3], and resistive-type [4] sensors have been developed for NOx detection. The limiting-current type detects the electric current generated by oxygen ions from NO when a voltage is applied, and the oxygen-ion current depends on the NO concentration. This sensor can detect NO in a wide range of concentration, and it has been used in cars with diesel engines. The mixed-potential type detects voltages produced by NOx. The resistive type detects current flowing through gas-detection materials under an applied voltage. Compared with the limiting-current and resistive types, the SiC-FET sensor does not need a current to flow through the gas detection materials, because the current flows on the surface of the SiC. Furthermore, since the SiC-FET amplifies the gas-sensing signals, it is more sensitive than the mixed-potential type [3].

[*]Yoshitaka Sasago, 1-280 Higashi-Koigakubo, Kokubunji, Tokyo 185–8601, Japan, +81-42-323-1111 & yoshitaka.sasago.tx@hitachi.com

The SiC-FET NOx sensor can detect NO in the mixture of gases in exhaust fumes when it is equipped with an ion-pumping device for preprocessing. The ion-pumping device can remove disturbance gases, and thereby, it enhances the selectivity of the sensor for NO gas [2, 5]. Moreover, without the on-chip ion-pumping device, the sensor can detect several gases such as oxygen (O_2), sulfur oxide (SO_2), carbon monoxide (CO), and ammonia (NH_3). These gases are also included in the exhaust from automobile engines. Instead of improving the sensor's selectivity by removing disturbance gases, we propose that it can be enhanced by using the temperature dependence of gas sensitivity. The high-temperature immunity of SiC-FETs suggests that they can be used for gas-sensing operation over a wide range of temperatures, and the sensitivity to each gas component of these devices varies depending on the sensing temperature.

In this study, we examined the temperature dependence of the gas responses of the SiC-FET sensor. We found that it can sensitively detect several different gases at high temperature. We conclude that a low-cost gas sensor with high sensitivity can be realized by taking advantage of semiconductor products.

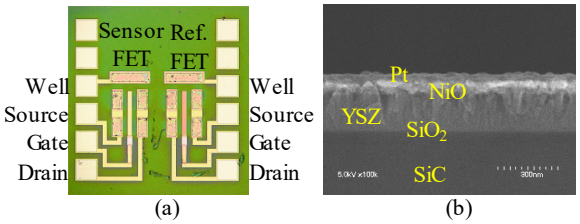


Figure 1: SiC-FET-type NOx sensors [1]: (a) photograph of SiC chip (2 mm × 2 mm), (b) gate layer of sensor FET.

2. Device Structure and Operation Principle

As shown in Fig. 1(a), the SiC-FET NOx sensor chip measuring 2 × 2 mm^2 includes a sensor FET, a reference FET, and platinum electrodes for the well, source, gate, and drain terminals on the edges on both sides [1]. Both FETs have a gate layer stack of yttria-stabilized zirconia (YSZ), nickel oxide (NiO), and platinum (Pt) on a gate oxide layer (Fig. 1 (b)). The platinum layer of the sensor FET is exposed to the atmosphere, while that of the reference FET is covered with dielectric films. The composition of YSZ is $(ZrO_2)_{0.97}(Y_2O_3)_{0.03}$. The process flow of the sensor chip is explained in Ref. [1]. Since NO, O_2, SO_2, CO, and NH_3 can only be sensed at high temperature, a heater should be placed near the sensor. The heater can be included in a module along with the sensor.

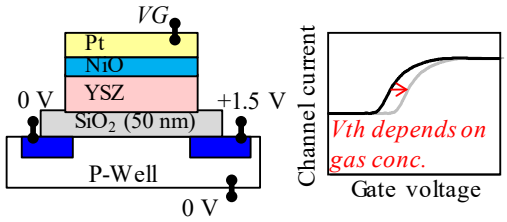


Figure 2: Device structure and operation principle of FET-type sensor [1].

The operation principle of the SiC-FET gas sensor is described in Ref. [1]. As shown in Fig. 2, the threshold voltage (*Vth*) of the sensor FET shifts by increasing the target gas concentration. For example, when the target gas is NO, *Vth* shifts in the positive direction as the concentration increases. However, *Vth* shifts in the negative direction for NH_3. The target gas concentration can be calculated from the *Vth* shift. A SiC-FET gas sensor was studied for hydrogen detection [6] and ammonia detection [7], but its gate structure was different from our device. The gate material can facilitate detection of target gases. The FET- and capacitor-type gas sensors are classified as work function-type sensors [8].

3. Gas Detection Material

FET sensors have been studied for detection of hydrogen [6, 9–10], ammonia [7], oxygen [11], NH_3 and H_2S [12], etc. The gate layers of these sensors consist of detection materials, which determine the gas detection characteristics of the sensors. Here, we examined the oxygen gas response of the sensor because our primary targets are oxygen-related gases. To select suitable gate-layer materials, we compared the oxygen gas response of the SiC-FET gas sensors with those of Si-FET gas sensors.

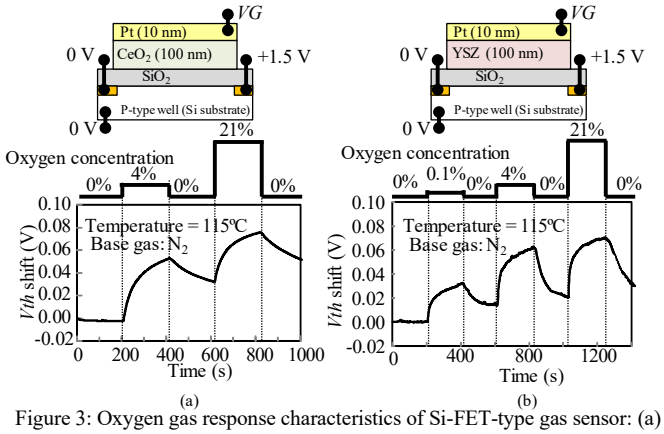


Figure 3: Oxygen gas response characteristics of Si-FET-type gas sensor: (a) platinum-ceria gate, (b) platinum-YSZ gate.

The oxygen response characteristics of the sensors with a platinum-YSZ gate and a platinum-ceria gate are compared in Fig. 3. The platinum, ceria, and YSZ films were deposited by sputtering. YSZ and ceria are oxygen ion conductors [13, 14], and platinum is a commonly used electrode for gas sensors [9–12]. Both sensors responded to oxygen gas; *Vth* shifted in the positive direction at 115°C, and the shifts depended on the oxygen concentration.

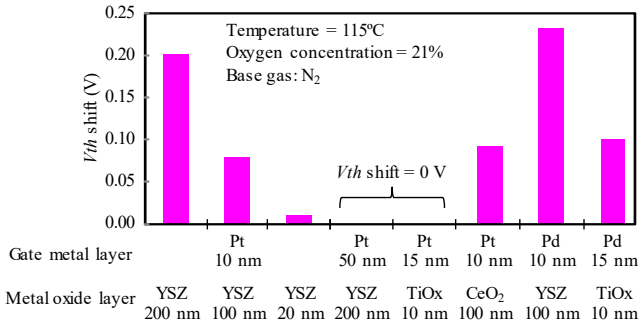


Figure 4: *Vth* shifts of Si-FET-type gas sensors with different gate layer materials.

The *Vth* shifts of different gate-layer materials are summarized in Fig. 4 for 21% oxygen at 115°C. The shifts depended on the thickness of the platinum layer, which is consistent with the results in Ref. [11]. The sensor with 10-nm-thick platinum responded to

oxygen gas, but the one with 50-nm-thick platinum did not. The thickness of YSZ also affected the *Vth* shift. The sensor with 10-nm-thick platinum and 200-nm-thick YSZ responded to oxygen gas. However, *Vth* shift decreased with decreasing YSZ thickness, and the sensor with 20-nm-thick YSZ hardly responded to oxygen gas. Platinum-titania gate FET responded to hydrogen gas, but not oxygen gas [9, 10]. The palladium-titania gate sensor responded to oxygen gas, but the platinum-titania gate sensor did not. The response of the sensor with 10-nm-thick palladium and 100-nm-thick YSZ was the largest of the test samples.

Among the oxide and metal materials in Fig. 4, platinum and YSZ are stable in terms of their chemistry and mechanics. Ceria is reduced in hydrogen atmosphere [14], and palladium is mechanically fragile because its volume expansion and contraction are large under a change in the hydrogen concentration. For these reasons, we selected platinum and YSZ for our SiC-FET gas sensor. The problem is that a thin-film platinum layer does not resist high temperatures over 500°C, while a sensor with a thick platinum layer (50 nm) does not respond to oxygen at 115°C, as shown in Fig. 4. However, a sensor thick platinum layer can do so at higher temperatures, as described below. We decided to use a 100-nm-thick platinum layer.

Although the sensor reported in Ref. [3, 15] is not an FET type, in that study it was shown that a catalytic metal oxide layer helps to enhance sensitivity for NOx [3, 15]. NiO and WO_3 are promising candidates for the catalytic layer, and one made of NiO shows excellent catalytic characteristics. Therefore, we decided to use a NiO layer in our sensor. In particular, NiO layers with thicknesses of 10 nm and 50 nm were formed between the platinum and YSZ layers. Moreover, composite material layers consisting of Pt and NiO were used instead of stacks of NiO and Pt.

The composition of the YSZ film was measured by X-ray photoelectron spectroscopy (XPS). It was the same as that of the sputtering target ($(ZrO_2)_{0.97}(Y_2O_3)_{0.03}$). No segregation of yttria was observed.

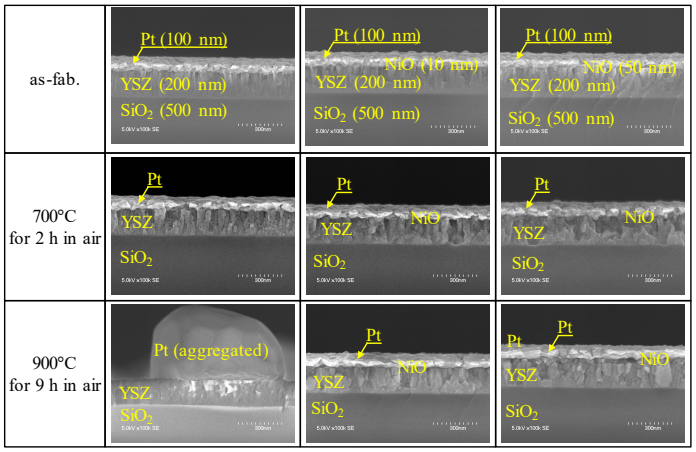


Figure 5: Heat resistance of sensor gates with 100-nm thick platinum. The photographs in the left and central columns are from Ref. [1].

The heat resistance of the gate stacks was examined by taking SEM micrographs. The results for the Pt-YSZ stacks and Pt-NiO-YSZ stacks are shown in Fig. 5. The platinum layer without the NiO layer became aggregated and disconnected at 900°C, but those with the 10-nm-thick and 50-nm-thick NiO layers remained continuous. The results for the stacks with composite material layers are shown in Fig. 6. The composite material layers also showed heat resistance up to 900°C. These results indicate that gate stacks with NiO could possibly achieve gas sensing at temperatures up to around 900°C.

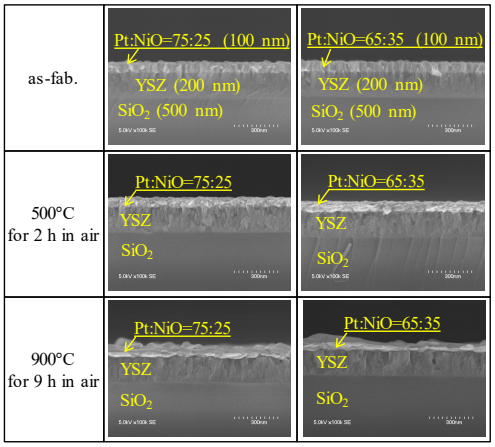


Figure 6: Heat resistance of sensor gates with Pt-NiO composite materials. The photographs in the left columns are from Ref. [1].

4. Gas-Sensing Characteristics

By virtue of the large band gap of SiC, SiC FETs are capable of on- and off-switching at over 500°C; thus, the *Vth* shift of the SiC-FET gas sensor can be detected [1]. On the other hand, switching at 500°C is impossible for Si FETs.

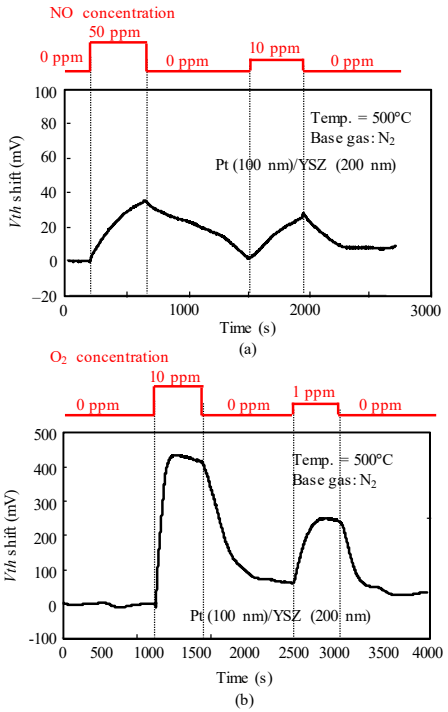


Figure 7: Gas response characteristics of the SiC-FET-type gas sensor with Pt-YSZ gate: (a) NO response, (b) O_2 response.

The time dependence of *Vth* was measured for the SiC-sensor FET. As Figs. 7 and 8 show, *Vth* depended on the concentrations of NO and O_2. Moreover, as shown in Fig. 7, the sensor with the Pt-YSZ gate responded to both NO and O_2, but its sensitivity to O_2

was higher. It clearly responded to 1 ppm O_2, with a *Vth* shift of about 250 mV, and it responded to 10 ppm NO with a *Vth* shift of about 25 mV. As well, the Pt-NiO-YSZ gate FET responded to both NO and O_2. The *Vth* shift for 1 ppm O_2 was about 200 mV, and that for 10 ppm NO was about 75 mV. The selectivity for NO and O_2 was thus modified by inclusion of the NiO layer.

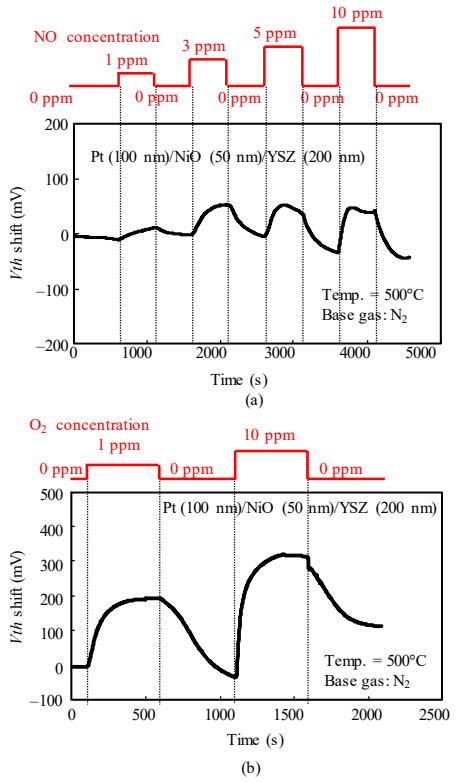


Figure 8: Gas response characteristics of the SiC-FET-type gas sensor with Pt-NiO-YSZ gate [1]: (a) NO response, (b) O_2 response.

We fabricated SiC-FET sensors on SiC wafers and measured the gate-length dependence of the FET characteristics of seven samples in air. In particular, *Vth* of the FET decreased smoothly as the gate length decreased; therefore, the sample variation was small. However, the gas response characteristics were measured in only one sample with a gate length of 100 nm for the sensor with the Pt-YSZ gate and the one with the Pt-NiO-YSZ gate. The sample-to-sample variation of gas responses thus remains an issue for future study.

Because of the sample variation problem, we decided to measure the responses of MOS-capacitor-type sensors with N-type SiC substrates instead for NO, O_2, SO_2, CO, and NH_3. While MOS capacitors are simpler than FETs, a change in the target gas concentration produces *Vth* shifts in these sensors like in FET sensors. Both FET- and capacitor-type gas sensors are classified as work function-type sensors [8].

As shown in Fig. 9, *Vth* of the Pt-YSZ gate capacitor increased as the concentration of NO gas increased, like the response of the FET sensor shown in Fig. 7(a). We measured the responses of a Pt-NiO-YSZ gate capacitor with 10-nm-thick NiO for NO gas at 700 and 800°C (Fig. 10) and those of a gate capacitor with 50-nm-thick NiO layer for NO gas at 600 and 800°C (Fig. 11). As Fig. 11 (a) shows, the sensor with 50-nm-thick NO layer responded to 0.5-ppm NO at 600°C, and the shift for 50 ppm NO was larger than 300 mV. Comparing Fig. 9 with Fig. 11 (a) suggests that the NiO layer appears to improve the NO sensitivity of the sensor. We also measured the NO responses of the capacitor-type sensor at 400°C, but the sensor hardly responded.

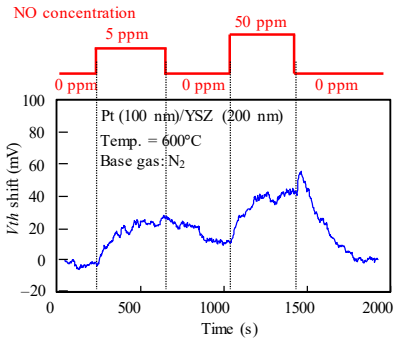


Figure 9: NO response characteristics of SiC-capacitor-type gas sensor with Pt-YSZ gate.

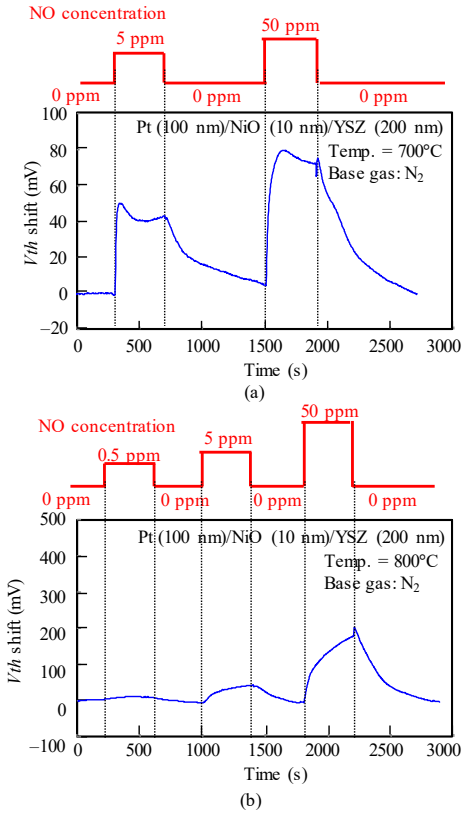


Figure 10: NO response characteristics of SiC-capacitor-type gas sensor with Pt-NiO (10 nm)-YSZ gate: (a) 700°C, (b) 800°C.

We measured the sensor responses of a Pt-YSZ gate capacitor for O_2 gas at 400 and 600°C and those of a Pt-NiO-YSZ gate capacitor with 10-nm-thick NiO for O_2 gas at 400, 600, and 800°C (Figs. 12 and 13). The sensors responded to an O_2 concentration of 0.1 ppm even at 400°C. The O_2 response could be detected at 400°C, while the NO response could not, because the catalytic reaction of the gate layers for O_2 is more active than for NO at low temperatures.

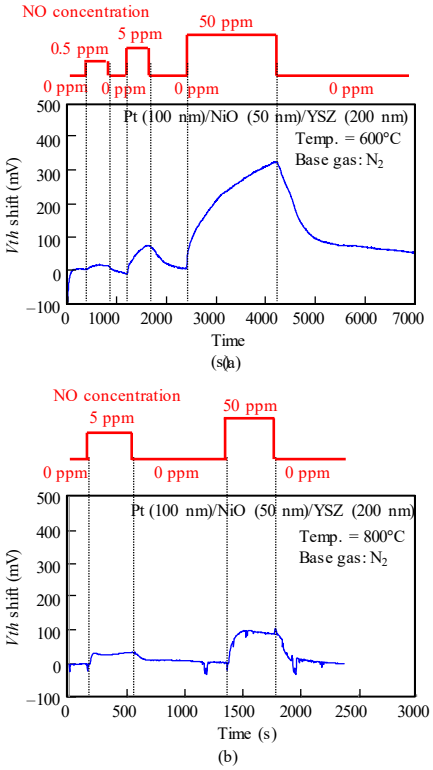


Figure 11: NO response characteristics of SiC-capacitor-type gas sensor with Pt-NiO (50 nm)-YSZ gate: (a) 600°C, (b) 800°C.

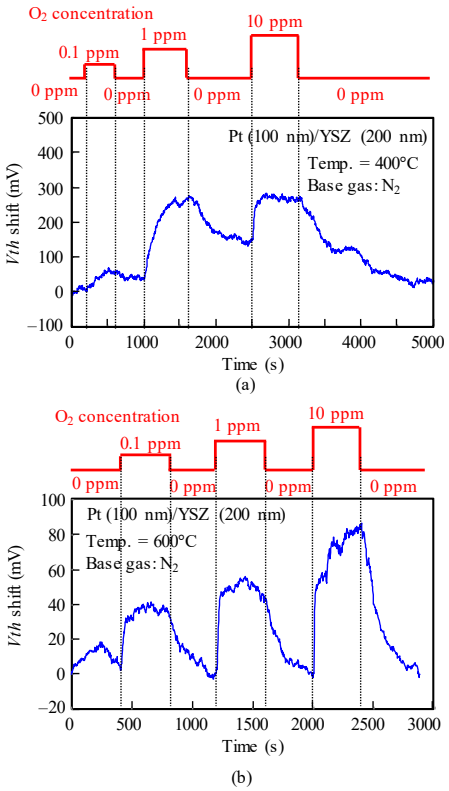


Figure 12: O_2 response characteristics of SiC-capacitor-type gas sensor with Pt-YSZ gate: (a) 400°C, (b) 600°C.

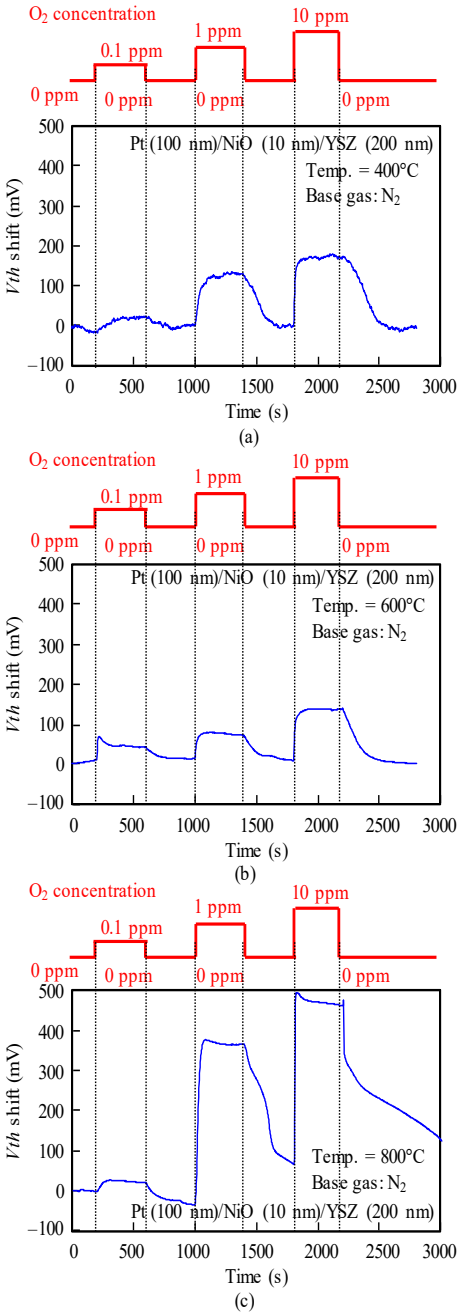


Figure 13: O_2 response characteristics of SiC-capacitor-type gas sensor with Pt-NiO (10 nm)-YSZ gate: (a) 400°C, (b) 600°C, (c) 800°C.

The responses of a Pt-YSZ gate capacitor were measured for NH_3 gas at 400, 600, and 700°C (Fig. 14), and those of a Pt-NiO-YSZ gate capacitor with 10-nm-thick NiO were measured for NH_3 gas at 400, 600, 700, and 800°C (Figs. 15 and 16). Although *Vth* shifted in the positive direction for NO and O_2 (Figs. 7–13), it shifted in the negative direction for NH_3. Note that a negative shift in *Vth* was previously observed in FET-type hydrogen sensors [9, 10]. Both sensors responded to NH_3 of 29 ppm concentration at 400°C. At 700°C, the response of the Pt-YSZ gate capacitor dropped rapidly, but that of a Pt-NiO-YSZ gate capacitor with 10-nm NiO decreased only slightly.

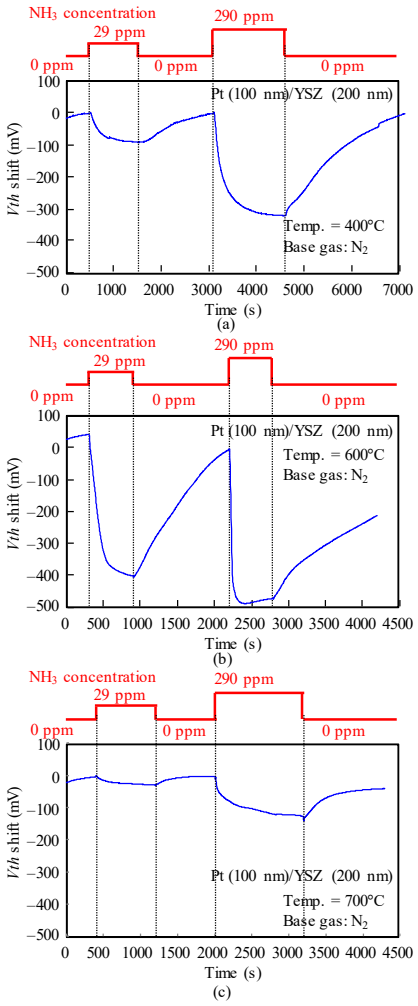


Figure 14: NH_3 response characteristics of SiC-capacitor-type gas sensor with Pt-YSZ gate: (a) 400°C, (b) 600°C, (c) 700°C.

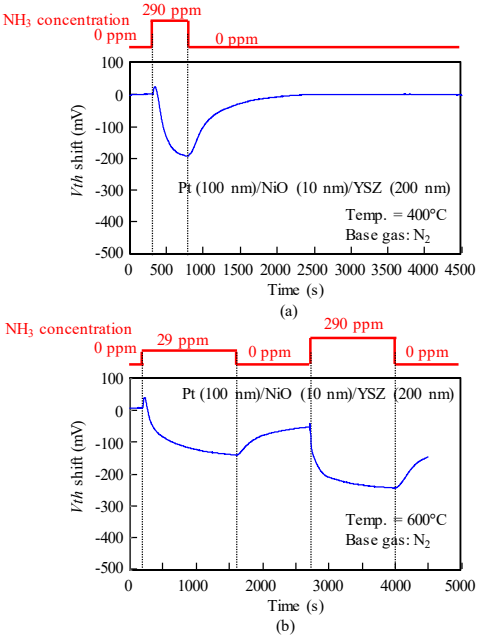


Figure 15: NH_3 response characteristics of SiC-capacitor-type gas sensor with Pt-NiO (10 nm)-YSZ gate: (a) 400°C, (b) 600°C..

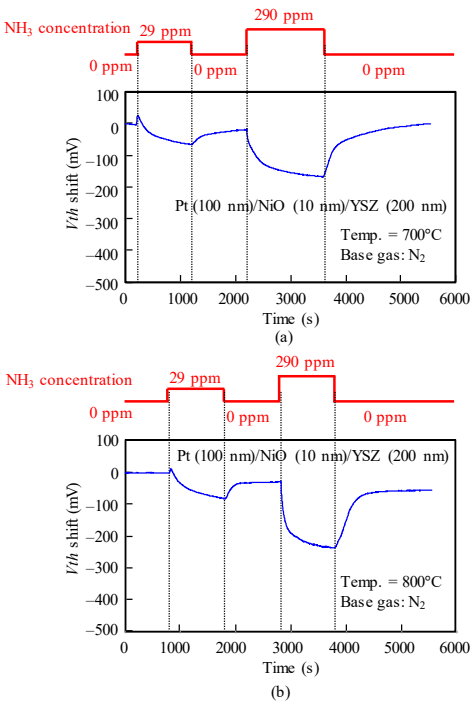


Figure 16: NH_3 response characteristics of SiC-capacitor-type gas sensor with Pt-NiO (10 nm)-YSZ gate: (a) 700°C, (b) 800°C.

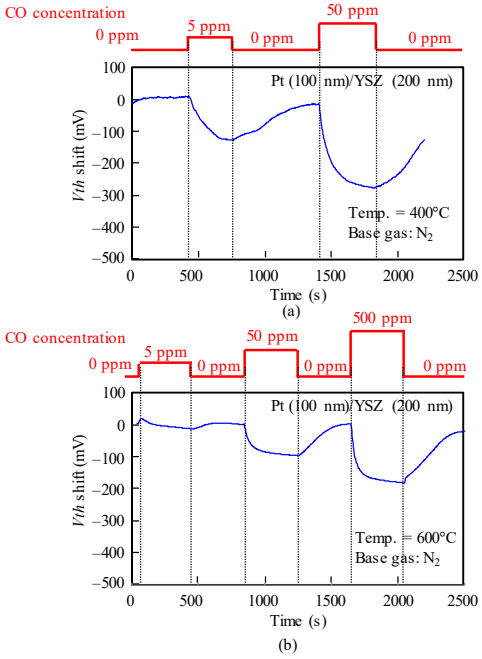


Figure 17: CO response characteristics of SiC-capacitor-type gas sensor with Pt-YSZ gate: (a) 400°C, (b) 600°C.

The sensor responses of the Pt-YSZ gate capacitor were measured for CO gas at 400 and 600°C (Fig. 17) and those of the Pt-NiO-YSZ gate capacitor with 10-nm-thick NiO were also measured for CO gas at 400, 600, and 700°C, as shown in Fig. 18. Similar to the responses for NH_3, *Vth* shifted in the negative direction for CO. The sensors responded to a CO concentration of a 5 ppm at 400°C. As the temperature increased, the sensor response of the capacitors decreased. The Pt-NiO-YSZ gate capacitor was measured at 800°C, but it did not respond to CO gas.

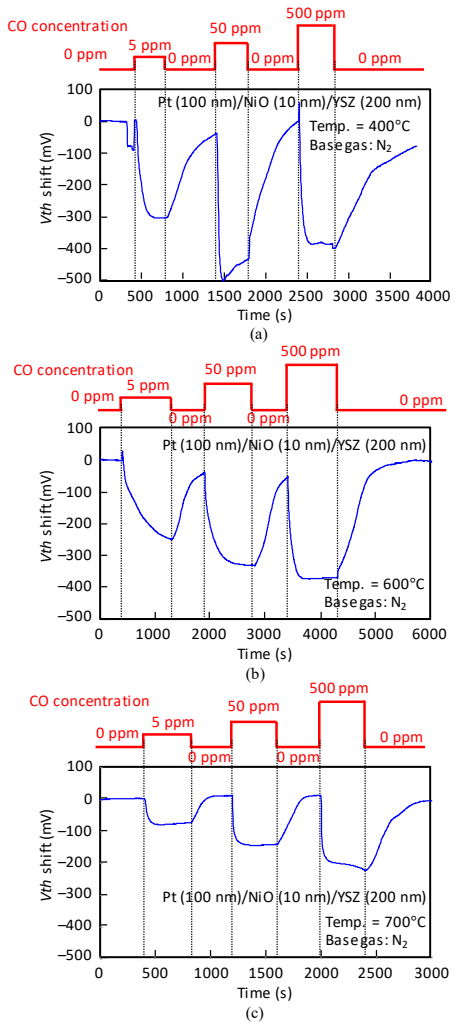


Figure 18: CO response characteristics of SiC-capacitor-type gas sensor with Pt-NiO (10 nm)-YSZ gate: (a) 400°C, (b) 600°C, (c) 700°C.

We measured the sensor responses of a Pt-NiO-YSZ gate capacitor with 50-nm-thick NiO for SO_2 gas at 600 and 800°C. As shown in Fig. 19, *Vth* shifted in the positive direction for SO_2 gas, as in the case of NO and O_2 gases. The sensor responded to a 5 ppm concentration of SO_2 at 600°C, and the response slightly decreased at 800°C.

The temperature dependence of the *Vth* shifts are summarized in Fig. 20 for NO (5 ppm), O_2 (1 ppm), NH_3 (29 ppm), CO (5 ppm), and SO_2 (5 ppm) gases. The results clearly show that the sensitivity depends on the measurement temperatures and the composition of the gate stack layer. Moreover, the measurement temperature can be controlled by the heater attached to the sensor module. Overall, the results indicate that measurement is possible over a range from 400 to 800°C, which is impossible for Si FET sensors.

The sensor responses of a Pt-YSZ gate capacitor and a Pt-NiO-YSZ gate capacitor were measured at room temperature, but no *Vth* shift was observed for NO (5 ppm), O_2 (1 ppm), NH_3 (29 ppm), CO (5 ppm), or SO_2 (5 ppm) gases. These results indicate our sensors have threshold temperatures for gas detection. Similar results are reported for NH_3 and H_2S in Ref. [12], in which the gas response was observed above 200°C for H_2S at a concentration of 400 ppm and above 110°C for NH_3 at a concentration of 500 ppm.

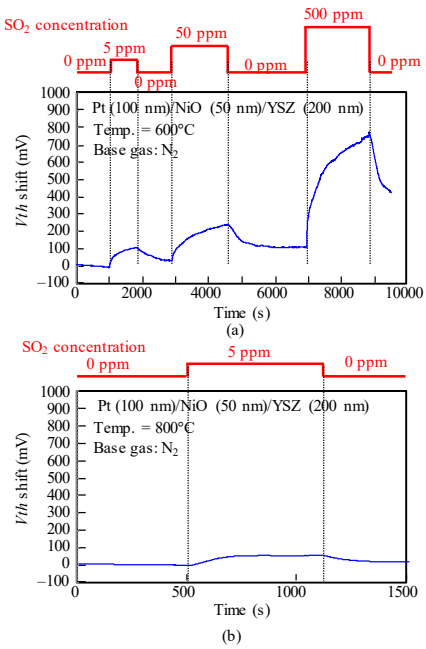


Figure 19: SO_2 response characteristics of SiC-capacitor-type gas sensor with Pt-NiO (50 nm)-YSZ gate: (a) 600°C, (b) 800°C.

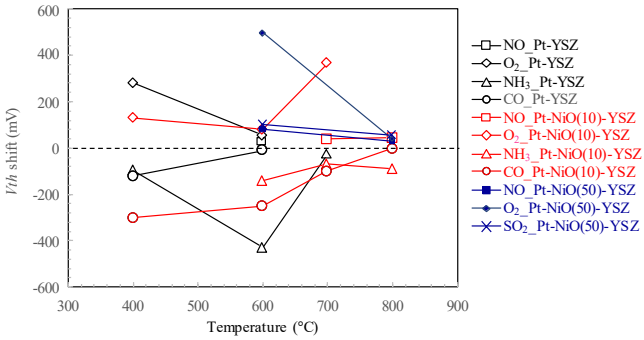


Figure 20: Temperature dependence of *Vth* shifts for NO (5 ppm), O_2 (1 ppm), NH_3 (29 ppm), CO (5 ppm), and SO_2 (5 ppm) gases.

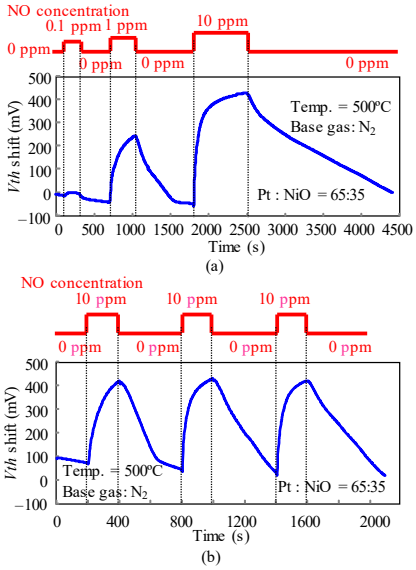


Figure 21: NO response characteristics of SiC-capacitor-type gas sensor with Pt-NiO-composite-material gate: (a) concentration dependence [1], (b) reproducibility.

To increase the sensitivity of the FET-type sensor, the three-phase boundary of the gate should be increased [16]. The nanoporous structure of the gate helps to improve the sensor sensitivity [9, 10]. The sensor with a NiO-Pt composite material gate showed NO detection with 0.1 ppm (Fig. 21 (a)). The reproducibility of the measurements was confirmed for 10 ppm NO, as shown in Fig. 21 (b). As mentioned above and as shown in Fig. 6, this sensor was very heat resistant.

5. Conclusion

We developed a SiC-FET-type gas sensors and SiC-capacitor-type gas sensors with a gate stack including YSZ, nickel oxide and platinum. These sensors are able to detect NO, O_2, NH_3, CO, and SO_2. To study the feasibility of the gas detection of each component in the gas mixture, we examined the temperature dependence of the gas response characteristics for each gas.

The nickel oxide layer enhances the heat resistance of the gate stack, and, thus, it extends the range of gas detection temperatures up to 800°C. The nickel oxide layer also helps enhance the sensitivity for NO detection. The sensors successfully detected NO at a concentration of 0.1 ppm. They were also able to respond to O_2 at a concentration of 0.1 ppm.

The gas response characteristics of capacitor-type sensors were measured or NO, O_2, NH_3, CO, and SO_2. The results show that *Vth* shifts in the positive direction for NO, O_2, and SO_2, but shifts in the negative direction for NH_3 and CO. The temperature dependence of the sensitivity was measured at room temperature and from 400 to 800°C. Gas responses were not observed at room temperature, and the sensitivity and the selectivity depended on the measurement temperature and the gate material of the sensors. This results indicate that the concentrations of each gas component can be determined using these dependences. In the future, we will study the sample variation of the gas responses characteristics and derive quantitative formulae for the concentration of each gas component.

In addition to the sensor with a gate stack of YSZ, NiO and Pt, the sensor with a NiO-Pt composite material gate was fabricated, and it was able to detect NO with a concentration of 0.1 ppm. This sensor was heat resistant, and its measurements were reproducible.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors gratefully acknowledge Dr. T. Sekiguchi and Dr. H. Kurata of Hitachi, Ltd. for all their encouragement. The authors also thank all the staff of the clean room of Hitachi's R&D group for the device fabrication. The authors thank Dr. R. Goto, Dr. H. Sano, Dr. T. Shimada, Mr. K. Kato, Mr. N. Ashizuka, and Mr. Y. Nozawa of Hitachi Metals, Ltd. for insightful discussion from the early stage of the research. The authors also thank Mr. Y. Nagai of TOMOE SHOKAI Co., LTD. for valuable assistance with the gas response experiments.

References

[1] Y. Sasago et al., "SiC-FET-type NOx Sensor for High-Temperature Exhaust Gas," 2018 IEEE International Electron Devices Meeting, 12.5.1–12.5.4.

[2] N. Kato, K. Nakagaki, and N. Ina, "Thick Film ZrO_2 NO_x Sensor," SAE Technical Paper 960334 (1996).

[3] N. Miura et al., "Improving NO_2 Sensitivity by Adding WO_3 during Processing of NiO Sensing-Electrode of Mixed-Potential-Type Zirconia-Based Sensor," J. Electrochem. Soc. 154, pp. 1246–1252 (2007).

[4] A. Groß et al., "Dual Mode NO_x Sensor: Measuring Both the Accumulated Amount and Instantaneous Level at Low Concentrations," Sensors 12, pp. 2831-2850 (2012).

[5] T. Inaba and T. Saji, "Low temperature operation of thin-film limiting-current type oxygen sensor using graded-composition layer electrodes," Sensors and Actuators B 129, pp. 874–880 (2008).

[6] R. Loloee et al., "Hydrogen monitoring for power plant applications using SiC sensors," Sensors and Actuators B 129, 200–210 (2008).

[7] H. Wingbrant et al., "Using a MISiC-FET Sensor for Detecting NH_3 in SCR Systems," IEEE Sensors Journal 5, pp. 1099–1105 (2005).

[8] T. Hübert et al., "Hydrogen sensors – A review," Sensors and Actuators B 157, pp. 329–352 (2011).

[9] T. Usagawa and Y. Kikuchi, "A Pt–Ti–O gate Si-metal-insulator-semiconductor field-effect transistor hydrogen gas sensor," J. Appl. Phys. 108, 074909 (2010).

[10] Y. Sasago et al., "FET-type hydrogen sensor with short response time and high drift immunity," Symp. on VLSI Technol. 2017, T106–T107.

[11] Y. Miyahara, K. Tsukada, and H. Miyagi, "Field-effect transistor using a solid electrolyte as a new oxygen sensor," J. Appl. Phys. 63, 2431 (1988).

[12] T. Usagawa and Y. Sasago, "Hydrogen-Related Gas Response from Pt-Ti-O Gate Si-MISFETs Gas Sensors," Proceedings of the 62nd Chemical Sensor Symposium 31, pp. 16-17 (2017).

[13] S. J. Skinner and J. A. Kilner, "Oxygen ion conductors," Materials Today 6, pp. 30–37 (2003).

[14] Y. Zhang et al., "Recent Progress on Advanced Materials for Solid-Oxide Fuel Cells Operating Below 500°C," Adv. Mater. 29, 1700132 (2017).

[15] A. Hemberga et al., "Effect of the thickness of reactively sputtered WO_3 submicron thin films used for NO_2 detection", Sensors and Actuators B 171, pp. 18-24 (2012).

[16] B. Wang et al., "Fabrication of Well-Ordered Three-Phase Boundary with Nanostructure Pore Array for Mixed Potential-Type Zirconia-Based NO_2 Sensor," ACS Appl. Mater. Interfaces, 8, pp. 16752–16760 (2016).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 159-168 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Testing Web Service Compositions: Approaches, Methodology and Automation

Dessislava Petrova-Antonova[1,*], Denitsa Manova[2], Sylvia Ilieva[1]

[1]*Software engineering, Sofia University, 1113, Bulgaria*

[2]*Rila Solutions, 1113, Bulgaria*

ARTICLE INFO

Article history:
Received: 05 November, 2019
Accepted: 20 December, 2019
Online: 22 January, 2020

Keywords:
Testing approaches
Testing methodology
Testing web service compositions

ABSTRACT

Web services give a new view of the web as the biggest, widely accepted and the most straightforward distributed software platform. Their composition into applications and business processes is still a complex, non-trivial task, requiring highly rational efforts not only from the software developers, but from the quality assurance specialists. The provision of web service compositions' quality brings a lot of challenges due to variability of difficulties at infrastructure, service and choreography levels and the need of different types of testing in unknown context and environment. A consolidated quality assurance methodology that advances the fundamental understanding of testing in terms of concepts, models, techniques, standards and automation is required. This methodology needs to enable effective exploration, comparison, evaluation and selection of testing approaches, platforms and tools.

This article proposes such a methodology and reviews the current testing approaches for single and composite web services from an objective, holistic perspective. The methodology is presented as an end-to-end testing procedure, in which activities are facilitated by a set of testing approaches, techniques and best practices. A concrete solution is recommended for actual implementation of each activity either through selection among the most appropriate and effective existing approaches or development of new approaches, mainly in case of critical issues such as dependencies analysis, partner web services' isolation and injection of faults. A common framework that integrates different testing tools automates the methodology. Its applicability, completeness, level of automation, and level of novelty is evaluated through testing of real testing scenarios.

1. Introduction

Web services provide a novel paradigm for interoperability of distributed applications. They act as collaborative agents to deliver advanced and high-quality functionality to the end users. That is why the web services are used by the successful enterprises to build flexible and fast connections with their partners, and thus reducing the cost and increasing the revenue. Such business interactions are possible through implementation of Web Service Compositions (WSCs) in a standardized, secure and interoperable manner. Although the WSCs provide a lot of benefits for the developers, they bring challenges to testing as well as to overall quality assurance process. First, its implementation is often based on web services, which are developed and deployed on diverse environments supported by different vendors. In case of legacy systems and components, the web services act as wrappers of functionality that is hard to be controlled and simulated in a testing environment. The orchestrated web services are not always available or could be undeployed by the provider, which leads to additional efforts for emulation of their behavior during testing. In addition, the emulation is required in case of payed web services to reduce the testing process's cost. Appropriate message data needs to be generated not only for the emulated web services to mimic their behavior, but for the composition as whole in order to be invoked in the testing environment. Achieving a high level of test coverage and determination of failure causes is difficult due to the missing programming code of the orchestrated services as well as the requirement for creating test cases based only on their public interfaces. Such interfaces are usually defined using the Web Service Description Language (WSDL). When a dynamic binding

*Dessislava Petrova-Antonova, 1113 Sofia, Bulgaria, 125 Tsarigradsko shose Blvd. Blok 2, +359887572094 & d.petrova@fmi.uni-sofia.bg

of web services is implemented, the emulation process becomes more complex, since advanced stubs (mockups) should be developed.

A lot of testing approaches and automated tools have been proposed to meet the above challenges. Unfortunately, the most of them cover only single testing activities such as analysis of testing paths, generation of test cases, web service mocking, injection of faults, etc. However, it is important to merge all testing activities in a general testing procedure supported by end-to-end methodology and thus to facilitate the whole testing process. This article is an extension of work originally presented in International Conference on the Quality of Information and Communications Technology (QUATIC'18) [1], which proposes such a methodology, called Testing As Service Software Architecture (TASSA). The methodology addresses the following major problems of WSC testing:

- Difficulties in emulation of web services deployed on heterogeneous platform and controlled by external providers;
- Missing or temporally unavailable web services and inability to detect the causes of failures;
- Lack of automation to test unanticipated behavior of web services and test case generation to reach a high level of test coverage.

Therefore, the benefit of the proposed methodology are as follows:

- Provision of end-to-end testing methodology for WSC;
- Recommendation of concrete approaches for actual implementation of each activity of the methodology;
- Development of completely new approaches to cover critical issues such as analysis of data dependencies, isolation of partner web services and injection of faults;
- Full automation in a single framework for testing WSCs, described with Business process execution language (WS-BPEL) [2].

The next two sections present the TASSA methodology and the testing approaches appropriate for its implementation. Section 4 provides a view of TASSA methodology's validation and Section 5 summarizes the validation results. Section 6 concludes the paper.

2. TASSA methodology

TASSA methodology supports functional, performance and security testing of WSCs and provide means for validation of their behavior if implementation changes are in place. It relies on a small number of artefacts, such as a BPEL file, input data, expected output data and test assertions. It consists of seven main activities, presented with a workflow diagram in Figure 1.

2.1. Prerequisites

W3C defines a set of standards for WSCs that are used for standard compliance validation of the WSCs [3]. If some inconsistencies are detected, a notification to perform the approapriate coreections are sent to the tester. The isolation of a partner web service can can follow two approaches: (1) invocation of mockup service that mimic the behavior of the partner web service or (2) generation of expected response that is expected from the partner web service. TASSA methodology adopts the second approach based on change of the communication channel between the WSC and its partner web service. This is realized using a so called "mediator service" allows fo injection of different types of faults in the communication channel.

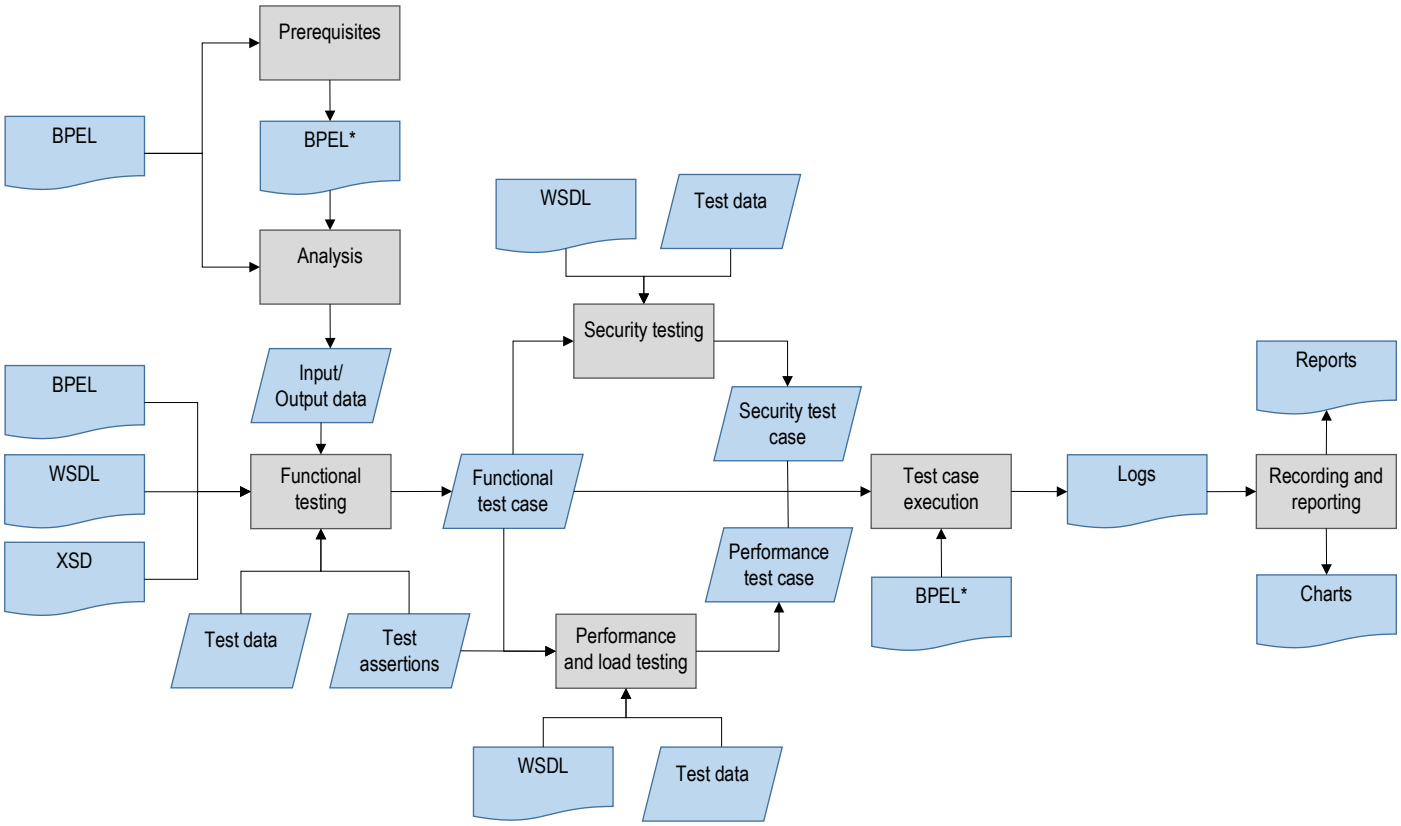


Figure 1: Methodology workflow

2.2. Analysis

In order to obtain all test paths, the WSC is usually represented using a specific formal model. TASSA methodology transforms the WSC in a BPEL Flow Graph (BFG) similarly to the approaches in [4] and [5]. First, all paths in the BFG are identified, next the unfeasible ones are filtered and the finally the feasible paths are used for generation of test cases.

2.3. Functional testing

The WSDL description of the WSC is the primary source for functional test case generation. The main components of a given test case are the request with input data and the expected response with output data. Similar test cases are a common practice for testing single web services. A modified approach that support tracing of the executed activities in the WSC is applied by the TASSA methodology.

The generation of test data is based on a widely adopted approach based on XML Schema Definition (XSD) of the WSC. It is further developed to support parametrization of test cases using variables, functions as well as connectivity with external data pools allowing for data-driven testing. The expected results from the execution of the WSC are a base for definition of test assertions. The execution order of the activities in the WSC is an important assertion, when the behavior of multiple web services is orchestrated. Additional assertions relevant to both single and composite web services are defined, based on a single value, a value type, a regular expression, an HTTP status code, a file type and size, etc.

2.4. Security testing

The request to the WSC is enriched with appropriate test artefacts for the purposes of security testing. The response from the WSC is checked whether it matches the security requirements, such as delivery of partial message, unauthorized access, unavailable partner web service, etc. At the message level, the security testing includes sending of "secure" request and checking the response sent by the web service. There are a lot of security mechanisms applicable to WSCs. The TASSA methodology proposes usage of WS-Security standard and QoS policies, which are based on enrichment of exchanged messages with a specific information based on WS-RM, WS-Addressing or MOTM. At transport level, the testing for security is related to testing of the authentication, including sending certificates like Kerberos, SSL, etc.

The TASSA methodology follows a new approach for the robustness testing. It transforms the WSC under test allowing for simulation of faults during its execution. The supported faults are an unavailable partner web service, a delay of the response from a partner web service, wrong structure or semantic of the response from a partner web service. The test assertions for security testing are defined in a similar way of functional testing.

2.5. Performance and load testing

The performance and load testing validate the non-functional characteristics of the WSC. Since, it requires a set of requests to be send in order to simulate parallel executions, a definition of virtual users is performed for single functional test case or a group of them for a given time interval. The threads are commonly used in the most cases. They are scheduled to define the way, on which the virtual users are managed. The schedule contains information about the time for starting and stopping of requests, execution order of requests and time interval between them, number of executions, etc.

The test assertions for performance testing are based on those for functional testing. Test assertions for time intervals, resource usage, file sizes and others characteristics are also defined. They provide insight for the system and network load, the hardware and software resources, etc. Additionally, a special assertion for checking the compliance with the Service Level Agreements (SLAs) is created.

2.6. Test case execution

Prior execution of functional test cases, the BPEL file of the WSC should be deployed on the application server and an instance of the WSC should be created. The request defined in the functional test case is sent to the WSC and its behavior is tested in a similar way to a single web service. The security testing requires to configure the security settings of the application server, where the BPEL file is deployed. The performance or load testing needs definition and management of threads according to the parameters of in the test cases. When a huge load needs to be simulated or hardware resources of the application server are limited, the execution of test cases is distributed among different machines.

2.7. Recording and reporting

The information obtained during the execution of the test cases is recorded in a log file. In order to allow for easily processing, the valuable outputs are collected in a test data storage. Apart from the directly obtained test results, additional metrics are recorded, which are related to the WSC itself, the application server, communication channels, the network and so on. These metrics are simple, but can be further aggregated for computing of more complex ones. In order to check whether the test is failed or passed, the expected output and actual one after test case execution are compared based on the test assertions. When the test is failed, the reason for the failure is collected for defect allocation as well as for a subsequent regression testing. The test reporting uses the data from the log files and the corresponding test data storage. The XML is the most commonly used data format for the test reports.

3. Testing Approaches

This section presents testing approaches that are appropriate for implementation of TASSA methodology. For each methodology's task several approaches are identified and the most appropriate ones are recommended. For particular tasks new approaches are developed and reasoning about their applicability is given.

3.1. Standard compliance checking

The preparation of WSC for testing includes three main activities, described in previews section, namely checking for compliance with standards, isolation of partner web services and injection of faults.

The compliance with the standards can be validated by the most of Integrated Development Environments (IDEs) such as

Eclipse BPEL Validator, Oracle ILINT and NetBeans XML Validate. It can be performed following three different techniques depending on the validation rules:

- Rules described with a programming code (e.g. code written with Java);
- Rules based on XML or OCL constraints;
- Rules defined with a model of the WSC description (e.g. UML model).

The proposed methodology relies on the Eclipse BPEL Validator. It processes the BPEL file as a DOM and for each node (assign, copy, from, to, etc.) creates rules for checking the constraints prescribed by the corresponding standard. The rules are written in Java and can be changed by the tester.

3.2. Isolation of partner web services

The replacement of partner web service with a stub (mockup) service is a technique widely used to remove external dependencies from the WSC [6,7]. An alternative approach is to generate a message that substitutes the response expected by the WSC from a partner web service [8,9,10]. The second approach requires the real communication with the partner web services to be recorded and appropriate interfaces and data to be manually defined.

For the purpose of TASSA methodology, a new approach is elaborated. It replaces the BPEL activities related to external dependencies (e.g. Invoke activities) with ones that are internal to the WSC and provide similar output to the original activities (e.g. Assign activities). Since the output of the simulated activities are known, the execution of the WSC can be controlled to follow a particular path. For this purpose, proper values for the variables in conditional branches should be defined. Identification of the sequence of branch conditions and the values for conditional variables is a heavily task but it is automated by the TASSA isolation tool.

3.3. Fault injection

The robustness of the WSC can be tested relying on methods for negative testing by causing faults in the WSC itself or in its communication with the partner web services. A classification of the faults that are typical for service-oriented systems can be found in [11], while the specific faults for WSC are described in [12]. In [13] authors propose an approach that injects faults in the partner web service and then perform coverage testing of the code responsible for recovery from failures (e.g. exception handlers). In [14] a fault injection technique is applied by exploring the behavior of the WSC regarding the injected faults and thus to assess the its fault tolerance capabilities.

The fault injection is the second new approach proposed by the TASSA methodology. It supports injection of delays, errors and other malfunctions in the message exchange between WSC and its partner web services. The BPEL file of the WSC is modified by replacing the call to a partner web service with a call to a proxy service. In case of faults, the behavior of the third-party services is out of the TASSA scope. Therefore, faults in the outgoing calls are not injected. The proposed approach requires the original partner web service to be called and then intervention in its response to be performed. For this reason, the proxy service calls the original partner web service and then using the information for the fault simulated performs the required action. When data errors need to be injected, the two options are available – to change the values of data fields, while keeping the structure of the message or to insert random errors in the message data, which could break the XML structure of the response. For the first case, a tool for random generation of data according to an XML schema is required.

3.4. Dependency analysis

The analysis of WSC is focused on states, transitions and related usage of BPEL activities that are part of complex interactions. The software systems support two kinds of interactions as follows:

- Data flow interactions, where the definition and values of the next data items depend on the definition and values of the previous ones.
- Control flow interactions along the execution path, in which the next executions depend on the previous ones;

The testing of the above interactions is covered by the data flow testing and control flow testing, respectively. Such techniques are applicable mainly to white-box testing. In the context of the service-oriented systems, there are additional interactions due to communication with external services, for which only the types of input and output data are known.

An overview of types of interactions, related dependencies and approaches for analysis is can be found in [15]. The current approaches for dependency analysis apply variety formal models. The most popular ones transform the WSC in a control flow graph [16,17], Petri net [18,19], a state graph [20] or an UML model [21,22].

For the purpose of TASSA methodology, a new approach for dependency analysis is developed. As was mentioned before, it transforms the WSC in a BFG in order to find all executable paths. The dependency analysis includes identification of the conditional statements together with their conditional variables and constants, which values guide the execution of the WSC on a certain path.

3.5. Test data generation

The test data generation is a complex task, since the programming code of the WSC itself and its partner web services is often unavailable. That is why the approaches for test data generation are “black box” based, meaning that they rely only on the scope and type of the input and output variables.

Since the communication in WSCs is performed through exchange of XML messages, a large group of test data generation techniques use WSDL descriptions of the web services in combination with their XSDs [23,24,25,26,27]. The XSDs determine the constraints over the data types that are useful for generation of both simple and complex test data.

The approach proposed for TASSA methodology is similar to those presented in [28,29,30,31], since it also processes XSDs. It produces XML instances from a given XML Schema available in the WSDL file or in the BPEL file. Thus, XML messages needed for communication of the WSC with its partner web services are generated. The proposed approach can be applied to functional and robustness testing of WSCs due to support of both correct and incorrect XML instances’ generation. The implemented algorithm supports generation of empty XML documents – XML instances, which structure follows a given XML Schema, but does not

contain concrete values for the XML elements. In addition, generation of XML instances populated with random valid and invalid data depending on the testing goal is provided. Finally, a manual specification of values for the XML elements or their loading from external file are supported.

3.6. Test assertions definition

The service-oriented systems require definition of test assertions for a large group of test artefacts such as file content, file properties, message content (both header and body), message properties, expressions (e.g. XPath), variable values, variable properties, time properties and sequences of activities' execution.

Durand and Green propose a test script model allowing for expression of predicates crossing over diverse inputs and to handle a richer spectrum of outcomes [32]. The test assertions are well-structured objects defined with identifier, source (normative conformance requirement), target (instance of a specific artifact type), pre-condition that should be fulfilled, predicate (logical expression over the target) and prescription level (level of imperativeness of the source requirement). The major benefit of such approach is that the target can be tested within its context. The test assertions are defined using XLST and XPath and can be applied to functional testing and non-functional testing.

The power of XPath functions and expressions is leverage by Schematron – a simple pattern and rule language suitable for document testing [33]. Since the handling of the namespace is difficult, when rules are written according to an XML schema, and when there is a namespace prefix in a value of an attribute or element, the language is primarily intended for XML instances. XMLUnit is a tool offering both Java and .NET interfaces, which also supports validation according to a given XML schema, assertion of the values of XPath queries or comparison of XML documents with the expected outcomes [34]. The approach behind the tool relies on JDOM to traverse the XML tree.

TASSA methodology adopts the approach proposed by Durand и Green. In addition to it, the XMLUnit is also utilized due to its high level of automation.

3.7. Functional test case generation

The approaches for generation of functional test cases can be divided in two groups depending on the type of testing technique – white box or black box. The black box techniques handle the WSC as a single service with known operations and message types defined in a WSDL document. On the other hand, the white box techniques explore the actual BPEL file to find information about activities, control flow and data flow. According to the formal model used for transformation of WSC to find executable paths and thus generate test cases, the approaches falls in three main groups: approaches based on CFG [16], approaches using Petri nets [18] and approaches adopting model checking technique such as Web Service Automata (WSA) [35], Stream X-machine (SXM) [36], UML 2.0 activity diagram [37], PROMELA [38], etc.

The solution for test case generation of TASSA methodology follows mixed approach – a combination of black box and with box techniques. It supports the following tasks: (1) determination of web service operations and BPEL variables; (2) generation of templates of SOAP request; (3) test assertions' definition at HTTP level, SOAP level and BPEL variable level; (4) execution of test cases by exchange of SOAP messages; (5) collection of test results for reporting and follow-up actions. The test cases are described and stored in an XML format, where the root element has two attributes – "name", corresponding to the test case name, and "template", indicating whether a data-driven testing can be performed. The other elements are a narrative description of the test case, a WSDL operation under test, an input data, a data source for the purpose of data-driven testing, and test assertions. The provided types of assertions verify the status code of the HTTP response and response time, validate the SOAP message in case of success or failure and can apply XPath expressions to a BPEL variable or to a SOAP body.

The test case provides information for the web service address, and the message parts like HTTP headers, authentication mechanisms if they are required, and data that is carried in the body of the SOAP request. The test result consists of HTTP headers, data that is carried in the body of the SOAP response, the BPEL variables' values and the execution time. In case of data-driven testing, parametrized test cases are generated, in which the actual data in the SOAP request body and the expected data in the SOAP response body are substituted with variables.

3.8. Performance testing

Since the web services are inherently concurrent, the concurrent issues continue to receive a huge interest. There are two primary solutions for implementation of concurrency models, namely event-based and thread-based. In thread-based applications each request is resolved in synchronous mode. Thus, each request is processed in a single thread. When the response is complete, the thread returns back to thread pool. If an external service is called to resolve the request, the thread has to wait for the response. In event-based applications each request is resolved in an asynchronous mode. The thread returns back to thread pool before the response is completed and it is ready to serve some other request. There is a single thread (or a small number of threads) that routes and manages all requests. If an external service is called to resolve the request, the process continue to execute without blocking and receives a response through a callback mechanism. The BPEL engines work following one of the two models. IBM WebSphere Process Server, ActiveBPEL Engine, Oracle BPEL Process Manager are thread-based, while BPWS4J and BPEL-Mora are event-based. A BPEL engine prototype with a high performance, proposed in [39], adopts and event-driven architecture and join patterns delivered by the Microsoft Concurrent Coordination Runtime (CCR). Apache ODE executes long-running business processes, described with BPEL. ODE's Java Concurrent Objects (JaCOb) framework ensures the runtime implementation of BPEL constructs at the instance level [40]. JaCOb supports a concurrency at application-level without relying on threads.

The type of concurrency model directly affects the performance of WSC. In addition to the Response time, other non-functional characteristics of web services related to their performance are Throughput, Availability, Accessibility and Successful rate. In the context of WSC, the performance depends on the orchestration of the partner web services, which can follow five patterns: sequence parallel, conditional, cyclic or discriminative execution. For each execution pattern, the quality factor of each partner service is calculated and then an aggregated factor for the whole WSC is delivered. The Response time, Availability, Reliability, Bandwidth and Cost are non-functional characteristics used for assessment of WSCs in [41]. A drawback of the proposed approach is that it considers only sequence

execution of partner web services. The approach presented in [42] covers all five execution patterns, but takes into account only three are non-functional characteristics (Response time, Reliability and Cost). In contrast to the approaches that apply aggregation method for quality assessment, there is a more realistic method, which accounts for the uncertainty of partner web services and estimate the quality of a WSC's probabilistically.

TASSA methodology proposes usage of combination of approaches. It is validated in the context of Apache ODE BPEL Engine. The generation of the virtual users us performed by external tool, which handles the WSC as a single service applying black box testing. The same number of requests following the same schedule are used for testing of WSC and its partner web services.

3.9. Security testing

The security testing of WSCs brings challenges due to their distributed nature and requires application of more than a single approach. Kabbani and Tilley consider the security of WSCs from three points of view: known threats and weaknesses, security mechanisms for web services and privacy, integrity and accessibility [43]. The Open Web Application Security Project (OWASP) works on improving the software security. It describes a set of security vulnerabilities, including those that affect the web services. It proposes tools and best practices for security assurance [44]. The OASIS consortium describes security quality factors such as Encryption, Authorization, Authentication, Non-Repudiation, Integrity, Availability, Privacy and Audit [45]. The WS-standards (WS- Security Policy, WS-Addressing, WS-SOAP Message Security, WS-Trust) prescribe mechanisms for secure communication between single web services, but they are applicable to WSCs too.

Several research works propose development of models for data access of single services [46,47,48]. The security assessment of WSCs is based on such models. An alternative approach is to use a web service mediator, called broker, which composes the web services against the specified security constraints. An optimal WSC is selected according the QoS properties [49]. The broker uses a repository to store security and quality properties of web services and updates these properties after each execution of WSC. The security certification of web services is a technique applicable to selection, discovery, and composition processes. It includes certificate issuing and management, certification-aware service discovery, certificate validation, and service consumption [50]. Biskup et. al propose a framework supporting a decentralized execution of WSCs, which guarantee the correctness and the security of the execution [51]. The approach is based on a container – an encrypted and authenticated data structure that is passed among the composed web services. An XML-based script language for definition of security policies of WSCs is presented in [52]. It is integrated to a platform for run-time monitoring and security analysis. Data sharing agreements are another solution for secure collaboration between parties [53]. They define the data sharing policies for authorizations, obligations and prohibitions. The policies indicate the authorized, obliged or denied actions together with their related data and subject.

TASSA methodology relies on validation according to the security standards, testing through fault injection, described in Section 3.3 and checking of the implemented security mechanisms. The quality factors of security in [45] are defined for single web services. The security tests check whether particular mechanisms and standards are applied or not. They can be performed on a transport level or on a message level. Since TASSA methodology are focused on WSCs, it proposes testing of the data exchange security, as it is prescribed in [53].

3.10. Recording and documentation of test results

The instrumentation is a popular method for monitoring and analysis of the execution of WSCs is [54,55,56]. It adds s specific code in the BPEL file, which collects execution information without changing the normal execution of the WSC. The collected information is sent to external auditing web service for processing. The approach is technology independent and is suitable for functional testing, since it allows monitoring of data flow and control flow. It can be applied to security testing, but it is not appropriate for performance testing due to invocation of external web service, which can slow down the performance.

The BPEL engine itself provides a detail information for the execution state of the WSC. For example, the ActiveBPEL Engine [57], Oracle BPEL Process Manager (Glassfish) [58], IBM WebSphere Process Server [59], Apache ODE [40] collect information about the Response time, the variables and activities as well as the invoked partner web services and the respective message exchange. Since the BPEL engines control the execution of WSC, the approaches and tools relying on their capabilities are technology dependent, but applicable to both functional and non-functional testing. SALMon is a service-oriented system for monitoring services in order to detect violations in service level agreements (SLAs) [60]. It is technology independent, including three types of components, namely Monitors, Analyzer and Decision maker. The Monitors consist of measuring instruments, the Analyzer checks the SLA rules, while the Decision maker performs corrections to satisfy SLA rules. An architecture for run-time monitoring of WSCs, described with BPEL, is proposed in [61]. The business logic of the composition and the monitoring functionality are clearly separated through implementation of two types of monitors – instance monitors and class monitors. The instance monitors are responsible for execution of a single instance of the composition, while the class monitors report aggregated information about all executed instances. Both monitors are specified through a specific language and are generated as Java programs, which can be deployed in the run-time environment of the monitor engine.

TASSA methodology adopts the approach relying on the BPEL engine for monitoring of WSC and subsequent reporting of the collected information. Such approach is suitable for all types of testing and provides information about assertions, variables, messages, control flow and communication with the partner web services. Its advantage is the lack of intervention in the WSC and early reporting of the test results.

4. Validation of TASSA methodology

The validation of the TASSA methodology is performed through its application for testing of a sample business process, called *Check Payment*. The testing is performed using NetBeans integrated development environment.

4.1. Functional testing

The definition of functional test cases includes the following steps:

1) Selection of BPEL file, containing description of the *Check Payment* business process;

2) Selection of test path from the data dependency tree, generated from the BPEL file;

3) Identification of variables and constants, from which the execution of the selected path depend on, using the data dependency analysis tool. The tool returns description of dependencies, shown in Listing 1;

Listing 1: Description of dependencies

```
This branch is ELSE
This branch is IF
This branch is IF Condition 0:
String representation:
(count($OrderPartnerOperationIn.Order_Input/ns0:state) = 0
or count($OrderPartnerOperationIn.Order_Input/ns0:city) = 0)
Depending on: OrderPartnerOperationIn
Located at:
\process\sequence[1]\if[1]\sequence[1]\if[1]\sequence[1]\if[1]
Condition 1: String representation: (true() =
$ValidateEMailOut.parameters/ns6:ValidateEMailResult)
Depending on: ValidateEMailOut
Located at: \process\sequence[1]\if[1]\sequence[1]\if[1]
Condition 2:
String representation:
($Validate_CreditCardOut.parameters/ns1:Validate_CreditCardResult = 1)
Depending on: Validate_CreditCardOut
Located at: \process\sequence[1]\if[1]
```

Listing 2: SOAP request

```
<SOAPenv:Envelope xsi:schemaLocation="..."
xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:SOAPenv="http://schemas.xmlSOAP.org/SOAP/envelope/"
xmlns:ord="..." xmlns:ord1="http://xml.netbeans.org/schema/OrderInfo">
<SOAPenv:Body>
  <ord:OrderPartnerOperation>
     <Order_Input>
       <ord1:firstname>?string?</ord1:firstname>
       <ord1:lastname>?string?</ord1:lastname>
       <ord1:email>?string?</ord1:email>
       <ord1:country>?string?</ord1:country>
       <!--Optional:-->
       <ord1:state>?string?</ord1:state>
       <!--Optional:-->
       <ord1:city>?string?</ord1:city>
       <ord1:postalCode>?string?</ord1:postalCode>
       <!--Optional:-->
       <ord1:address>?string?</ord1:address>
       <!--Optional:-->
       <ord1:phone>?string?</ord1:phone>
       <ord1:creditCardNumber>?string?</ord1:creditCardNumber>
       <ord1:total>?1.051732E7?</ord1:total>
       <ord1:currencyCode>?string?</ord1:currencyCode>
       <!--Optional:-->
       <ord1:errorCode>?string?</ord1:errorCode>
    </Order_Input>
  </ord:OrderPartnerOperation>
</SOAPenv:Body>
</SOAPenv:Envelope>
```

Listing 3: Input test data

```
<SOAPenv:Envelope xsi:schemaLocation="..."
xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:SOAPenv="http://schemas.xmlSOAP.org/SOAP/envelope/"
xmlns:ord="..." xmlns:ord1="http://xml.netbeans.org/schema/OrderInfo">
<SOAPenv:Body>
  <ord:OrderPartnerOperation>
     <Order_Input>
       <ord1:firstname>John</ord1: firstname>
       <ord1:lastname>Doe</ord1:lastname>
       <ord1:email>john@gmail.com</ord1:email>
       <ord1:country>USA</ord1:country>
       <ord1:postalCode>77590</ord1:postalCode>
       <!--Optional:-->
       <ord1:address>St. John 114</ord1:address>
       <!--Optional:-->
       <ord1:phone>00359111222</ord1:phone>
       <ord1:creditCardNumber>41111111</ord1:creditCardNumber>
       <ord1:total>40</ord1:total>
       <ord1:currencyCode>USD</ord1:currencyCode>
    </Order_Input>
  </ord:OrderPartnerOperation>
</SOAPenv:Body>
</SOAPenv:Envelope>
```

4) Generation of abstract test case for testing of business process's operation *OrderPartnerOperation*. The SOAP request to the partner web service is shown in Listing 2;

5) Creation of executable test cases from the abstract one by manual or automated specification of values for the input variables, considering the identified constraints on step 3. The expected results and assertions are defined using the test case generation tool. In order to execute the path that contains the activity *CityAssign*, the credit card number and the client email should be valid, while the city and the state may be omitted. The input data that satisfy these conditions as well as standard data types in the XSD schema of the business process are presented in Listing 3;

The following assertions are defined:

- HTTP status code – checks the HTTP status code of the response. The expected value is 200;
- XPath equals – the BPEL variable *ValidateEMailOut* should contain XPath expression *message/part/ValidateEMailResponse/ValidateEMailRe* with value *True*.
- XPath equals – the BPEL variable *Validate_CreditCardOut* should contain XPath expression *message/part/Validate_CreditCardResponse/Validate_CreditCardResult* with value *1*.
- Not XPath exists – the BPEL variable *OrderPartnerOperationIn* should not contain XPath expression *message/part/city*.
- Contains – the response is expected to contain the regular expression *<ord1:city>.*</ord1:city>*;
- Response time – the maximum value of the Response time is expected to be 5 000 ms.

6) The test cases are executed and the results are recorded and analyzed.

Table 1: Output from functional testing

No	Assertion	Result
1	HTTP status code	200
2	XPath equals on *ValidateEMailOut* variable	True
3	XPath equals on *Validate_CreditCardOut* variable	1
4	Not XPath exists on *OrderPartnerOperationIn* variable	XPath not found
5	Contains (case sensitive: "false", is regex: "true")	Value found
6	Response time	2562 ms

4.2. Isolation of partner web service

Since the partner web service *addresslookup* allows limited invocations over time, it is isolated from the business process by performing the following steps:

1) Selection of BPEL file, containing description of the *Check Payment* business process.

2) Selection of *addresslookup* web service for isolation;

3) Formal description of isolation, shown in Listing 4. The partner web service is executed though invoke activity *PostCodeInvoke*. The result from invocation is replaced with a string constant "TEXAS CITY".

4) Isolation of the partner service through transformation of the business process. Figure 2 shows the transformed part

of the business process, where the *Assign1* activity replaces the original *PostCodeInvoke* invoke activity.

Listing 4: Formal description of isolation

```
C: /CheckPayment/src/CheckPayment.bpel
---
process//invoke[@name='PostCodeInvoke']
$OrderPartnerOperationOut.Order_Output/ns0:city='TEXAS CITY'
$OrderPartnerOperationOut.Order_Output/ns0:state='TX'
---
```

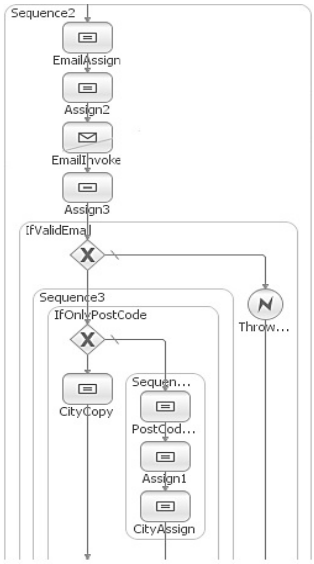


Figure 2: Isolation of partner web service

5) Execution of the transformed business process without invocation of *addresslookup* web service.

The output from isolation of partner web service is the same as one presented in Table 1. The partner web service *addresslookup* is successfully isolated without breaking of the normal execution of the business process.

4.3. Fault injection

The robustness testing of the business process is performed through the following steps:

1) Selection of BPEL file, containing description of the *Check Payment* business process.

2) Formal description of the fault that will be simulated, shown in Listing 5 – a delay in the communication channel between the business process and its partner web service *ValidatorDemo*.

Listing 5: Formal description of fault injection

```
C:/CheckPayment/src/CheckPayment.bpel
---
+proxy
wait=20
error_ratio=0
http://mathertel.de/AJAXEngine/S03 AJAXControls/ValidatorDemo.asmx
$ValidateEMailIn.parameters=$ValidateEMailOut.parameters
---
```

3) Transformation of the business process according to the formal description in step 3. The invocation of partner web service *ValidatorDemo* is replace with an invocation of proxy web service, called *PoxyInvoke*, which injects fault.

4) Execution of the transformed business process. The *ProxyInvoke* web service simulates a delay of 20 sec of the response from the partner web service.

The output from fault injection is the same as one presented in Table 1, except the value for the Response time. Due to delay of response from the partner web service *ValidatorDemo*, the actual Response time of the business process is 23718 ms.

5. Summary of methodology validation

TASSA methodology is evaluated based on the following assessment criteria:

- ***Applicability*** to SOA-based systems;
- ***Completeness*** of testing related to both WSCs and its components;
- ***Level of automation*** of the whole testing process;
- ***Level of novelty***.

More details about the results form evaluation of TASSA methodology can be found in [62,63,64,65]. They prove that the TASSA methodology provides solution of the *most significant problems* related to the WSC testing: (1) *Inability to instrument web services* that are not under control of the developer; (2) *Delay of testing* due to parallel development of the application components that have dependencies between each other; (3) *Lack automation of testing* for situations when the communication channel is interrupted or there is a noise in it and (4) *Difficulties during identification of the root causes* of failures and lack of a specific analysis approach dealing with different technology platforms.

The validation results shows that the isolation of partner web services can be applied to both synchronous and asynchronous WSCs. The isolation of the external dependencies and the injection of faults keep the original behavior of the WSC. In addition, the proposed approach for fault injection successfully simulate different types of failures. The identification and tracking the execution paths of WSC is supported by TASSA methodology through the approach for dependency analysis. This approach successfully identifies the conditional activities in the BPEL file and the variables, which values determine the execution on a given path.

The evaluation of TASSA methodology according to the above criteria proves its feasibility and effectiveness. The TASSA methodology proposes ***a novelty solution for testing SOA-based applications***. Its level of novelty is directly related to the maximum level of novelty of the adopted approaches. Therefore, the level of novelty of TASSA methodology is 3 from total 3. It delivers ***a complete solution***. Its completeness is calculated as a sum of completeness of the adopted approaches. The completeness of TASSA methodology is 11 from total 12. The TASSA methodology provides ***a high level of automation***. Its level of automation is computed as a sum of level of automation of the approaches that are proposed for its implementation. This metric varies due to possibility for combination of different approaches to achieve a specific testing goal. Since the testing approaches share test data, the manual activities related to its generation on the latest testing stages are minimized. The average level of automation of TASSA methodology is 7 from total 12. It is fully applicable to testing of WSCs, defined with WS-BPEL. The ***applicability*** of the TASSA methodology is directly related to the minimum applicability of the approaches that are proposed for its implementation. It depends on the concrete approaches used to test a given WSC. When all activities of TASSA methodology are

performed, the applicability is calculated as 1. When the activities related to isolation of partner web services and dependency analysis are not performed, the applicability is calculated as 3. In this case the TASSA methodology can be applied to a number of service-based applications.

Table 2 Summarizes the evaluation results. The first fifth rolls are related to the proposed testing approaches, while the last one shows the results for the TASSA methodology as a whole.

Table 2: Summary evaluation of TASSA methodology

Approach/ methodology	Novelty	Completeness	Automation	Applicability	Total
Isolation of dependencies	100%	100%	66%	25%	73%
Fault injection	66%	100%	66%	75%	77%
Dependencies analysis	100%	33-66%	33-66	25%	48-73%
Functional test case generation	33%	100%	100%	100%	66%
Test data generation	66%	66-100%	66%	75%	68-77%
Overall methodology	100%	80-93%	33-100%	25-100%	66-77%

6. Conclusion

The present work addresses the challenges of testing WSCs by proposing a novel methodology for testing, called TASSA. The methodology consists of the following main activities: (1) Checking for compliance with standards and isolation of partner web services; (2) Definition of testing goal and dependency analysis of the WSC to obtain testing paths; (3) Definition of functional test cases; (4) Definition of test cases for performance and load testing; (5) Definition of test cases for security testing; (6) Test case execution; and (7) Recording and reporting of test results. The optional execution of some activities enables focusing of the testing process on the specific user requirements and performing a step-by-step testing. For example, preparation activities could be performed at the beginning of business process development, including checking for compliance with standards. On the next step, when the partner web services are identified and are available, a functional testing could be performed. Finally, when the development of the business process under test is completed, non-functional testing could be conducted. A significant benefit of the TASSA methodology is that it deals with a small set of test artefacts – BPEL file, test data and test assertions. When the methodology is applied to the functional testing, a high level of automation can be achieved, in which only the BPEL file should be provided.

A lot of approaches are explored to identify the most suitable ones for implementation of the activities of TASSA methodology. New approaches are proposed for the most critical issues such as fault injection, isolation of partner services and dependency analysis. For other activities current approaches are adopted and are extended if needed. Two popular approaches for functional test case definition and test data generation are extended. The automation of TASSA methodology is shown onto a sample, yet realistic case study as a proof-of-concept. The complete validation is performed using different business processes in terms of number of activities, number of partner web services and type of communication – synchronous or asynchronous. The novelty, automation and applicability of TASSA methodology to testing serviced-based applications are evaluated regarding clearly defined assessment criteria. The results shows that it fully support end-to-end testing of WSCs covering all required testing activities.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research work has been supported by GATE “Big Data for Smart Society” project, funded by the European Union's Horizon 2020 WIDESPREAD-2018-2020 TEAMING Phase 2 programme under grant agreement no. 857155 and Big4Smart “Big Data Innovative Solutions for Smart Cities”, funded by the Bulgarian National Science fund, under agreement no. DN12/9 and agreement no. DN 02/11.

References

[1] Petrova-Antonova, S. Ilieva, D. Manova, “TASSA Methodology: End-to-end Testing of Web Service Compositions” in 11th International Conference on the Quality of Information and Communications Technology (QIATIC’2018), September 4-7, 2018, 264-267, Electronic ISBN: 978-1-5386-5841-3. https://doi.org/10.1109/QUATIC.2018.00046

[2] OASIS Web Services Business Process Execution Language, https://www.oasis-open.org/, last accessed 2019/11/04

[3] W3C., https://www.w3.org/standards, last accessed 2019/11/04

[4] T. Lertphumpanya, T. Senivongse, “Basis path test suite and testing process for WS-BPEL” WSEAS Transactions on Computers, 7(5), 483-496, 2008.

[5] Y. Yuan, Li, Z., W. Sun, “A graph-search based approach to BPEL4WS test generation” in IEEE Int. Conf. on Software Engineering Advances, 2006. https://doi.org/10.1109/ICSEA.2006.261270

[6] IBM - Rational Test Virtualization Server. (2016, January 1). http://www.ibm.com/software/products/en/rtvs, last accessed 2019/11/04

[7] Web Services API Mocking Overview | SOAP Mocking. (n.d.). https://www.soapui.org/soap-mocking/service-mocking-overview.html. last accessed 2019/11/04

[8] CA Service Virtualization - CA Technologies. (n.d.). https://www.ca.com/us/products/ca-service-virtualization.html, last accessed 2019/11/04

[9] R. Fletcher, “Betamax - Record & replay HTTP traffic”, http://betamax.software/, 2011, last accessed 2019/11/04

[10] Service Virtualization: Application & Data Simulation Software | Hewlett Packard Enterprise. (n.d.). http://www8.hp.com/us/en/software-solutions/service-virtualization/, last accessed 2019/11/04

[11] S. Bruning, S. Weissleder, M. Malek, “A fault taxonomy for service-oriented architecture” in IEEE 10th High Assurance Systems Engineering Symposium, 2007, 367-368. https://doi.org/10.1109/HASE.2007.46

[12] K. M. Chan, J. Bishop, J. Steyn, L. Baresi, S. Guinea, “A fault taxonomy for web service composition” in International Conference on Service-Oriented Computing, 2007, Lecture Notes in Computer Science, vol 4907. Springer, Berlin, Heidelberg, 363-375. https://doi.org/10.1007/978-3-540-93851-4_36

[13] C. Fu, B. G. Ryder, A. Milanova, D. Wonnacott, “Testing of java web services for robustness” ACM SIGSOFT Software Engineering Notes, 29(4), 2004, 23-34. https://doi.org/10.1145/1007512.1007516

[14] M. G. Fugini, B. Pernici, F. Ramoni, “Quality analysis of composed services through fault injection” Inf Syst Front (2009) 11: 227. https://doi.org/10.1007/s10796-008-9086-3.

[15] P. Kumar, Ratneshwer, “A Review on Dependency Analysis of SOA based System” in IEEE Fifth International Conference on Recent Trends in Information, Telecommunication and Computing, 2014, 69-81. https://doi.org/02.ITC.2014.5.11

[16] J. Yan, Z. Li, Y. Yuan, W. Sun, J. Zhang” Bpel4ws unit testing: Test case generation using a concurrent path analysis approach” In IEEE 17th International Symposium on Software Reliability Engineering, 2006, 75–84. https://doi.org/10.1109/ISSRE.2006.16

[17] Z. J. Li, H. F. Tan, H. H. Liu, J. Zhu, N. M. Mitsumori, “Business-process-driven gray-box SOA testing” IBM Systems Journal, 47(3), 2008, 457-472. https://doi.org/10.1147/sj.473.0457

[18] W. L. Dong, H. Yu, Y. B. Zhang, “Testing BPEL-based web service composition using high-level petri nets” in 10th IEEE International Enterprise Distributed Object Computing Conference, 2006. 441-444. https://doi.org/10.1109/EDOC.2006.59

[19] W. M. Van der Aalst, M. Dumas, C. Ouyang, A. Rozinat, E. Verbeek, "Conformance checking of service behavior" ACM Transactions on Internet Technology, 8(3), 13, 1-30, 2008. https://doi.org/10.1145/1361186.1361189
[20] M. Daghaghzadeh, M. Babamir, "An ABC based approach to test case generation for BPEL processes" in IEEE 3th International eConference on Computer and Knowledge Engineering, 2013, 272-277. https://doi.org/10.1109/ICCKE.2013.6682849
[21] H. Cao, S. Ying, D. Du "Towards model-based verification of BPEL with model checking" In Sixth IEEE International Conference on Computer and Information Technology, 2006, Seoul, 190-190. https://doi.org/10.1109/CIT.2006.185
[22] Z. Guangquan, R. Mei, Z. Jun, "A business process of web services testing method based on UML2. 0 activity diagram" in IEEE Workshop on Intelligent Information Technology Application, Zhang Jiajie, 2007, pp. 59-65. https://doi.org/10.1109/IITA.2007.83
[23] C. Bartolini, A Bertolino, E. Marchetti, A. Polini, "WS-TAXI: A WSDL-based testing tool for web services" In IEEE International Conference on Software Testing Verification and Validation, 2009, 326-335. https://doi.org/10.1109/ICST.2009.28
[24] C. Ma, C. Du, T. Zhang, F. Hu, X. Cai, "WSDL-based automated test data generation for web service" In International Conference on Computer Science and Software Engineering, 2008, 731-737. https://doi.org/10.1109/CSSE.2008.790
[25] H. M. Sneed, S. Huang "WSDLTest-a tool for testing web services" in Eighth IEEE International Symposium on Web Site Evolution, 2006, pp. 14-21. https://doi.org/10.1109/WSE.2006.24
[26] J. Offutt, W. Xu "Generating test cases for web services using data perturbation" ACM SIGSOFT Software Engineering Notes, 29(5), 2004, 1-https://doi.org/10. 10.1145/1022494.1022529
[27] X. Bai, W. Dong, W. T. Tsai, Y. Chen, "WSDL-based automatic test case generation for web services testing" in IEEE International Workshop on Service-Oriented System Engineering, 2005, 207-212. https://doi.org/10.1109/SOSE.2005.43
[28] DataGen - XML Test Data Generation Tool (2012). http://iwm.uni-koblenz.de/datagen/index.html, last accessed 2019/11/04
[29] Eclipse XML Editor Help - Eclipse Platform. (n.d.). http://help.eclipse.org/neon/index.jsp?topic=%2Forg.eclipse.wst.xmleditor.doc.user%2Ftopics%2Fcwxmledt.html, last accessed 2019/11/04
[30] M. Sorens (2009). Taking XML Validation to the Next Level: Introducing CAM. DevX. http://www.devx.com/xml/Article/41066, last accessed 2019/11/04
[31] V. Bergmann, Databene Benerator 0.7.6, manual, 2012.
[32] J. Durand, S. D. Green, S. Kulvatunyou, T. Rutt, "Test Assertions on steroids for XML artifacts" In Balisage: The Markup Conference, vol. 3, 2009.
[33] J. Tennison, "Validating XML with Schematron", Beginning XSLT, 2004, Apress.
[34] T. Bacon, J. Martin, "XMLUnit - Unit Testing XML for Java and .NET", http://www.xmlunit.org/, last accessed 2019/11/04
[35] Y. Zheng, J. Zhou, P. Krause "An Automatic Test Case Generation Framework for Web Services" Journal of Software, 2(3), 20017, 64-77. http://epubs.surrey.ac.uk/id/eprint/1975
[36] Ch. Ma, J. Wu, Tao Zh., Y. Zhang, X. Cai "Automatic Test Case Generation for BPEL Using Stream X-Machine" International Journal of u- and e-Service, Science and Technology, ISSN: 2005-4246, 2008, 27-36.
[37] Q. Yuan, J. Wu, Ch. Liu, L. Zhang "A model driven approach toward business process test case generation" in 10th International Symposium on Web Site Evolution, 2008. https://doi.org/10.1109/WSE.2008.4655394
[38] J. García-Fanjul, J. Tuya, Cl. de la Riva "Generating Test Cases Specifications for BPEL Compositions of Web Services Using SPIN" in International Workshop on Web Services Modelling and Testing, 2006, 83-94.
[39] W. Chen, J. Wei, G. Wu, X. Qiao' Developing a Concurrent Service Orchestration Engine Based on Event-Driven Architecture" Meersman R., Tari Z. (eds) On the Move to Meaningful Internet Systems: OTM 2008. OTM 2008. Lecture Notes in Computer Science, vol 5331. Springer, Berlin, Heidelberg, 2008, pp. 61-68.
[40] Apache ODE Team. Apache ODE BPEL Engine - Architectural Overview, 2013. http://ode.apache.org/developerguide/architectural-overview.html, last accessed 2019/11/04
[41] B. Li, X. Y. Tang, J. Lv, "The research and implementation of services discovery agent in web services composition framework". in IEEE International Conference on Machine Learning and Cybernetics, 2005, vol. 1, 78-84.
[42] S. Y. Hwang, H. Wang, J. Tang, J. Srivastava, "A probabilistic approach to modeling and estimating the QoS of web-services-based workflows", Information Sciences, 177(23), 2007, 5484-5503.
[43] N. Kabbani, S. Tilley, "Evaluating the capabilities of SOA security testing tools" In IEEE International Systems Conference, 2011, 129-134. https://doi.org/10.1109/SYSCON.2011.5929125
[44] R. Groenboom, R. Jaamour, "Securing Web Services" in OWASP Europe Conference, Leuven, Belgium, 2006
[45] OASIS Web Services Quality Factors Version 1.0. OASIS Committee Specification 01. http://docs.oasis-open.org/wsqm/WS-Quality-Factors/v1.0/cs01/WS-Quality-Factors-v1.0-cs01.html, last accessed 2019/11/04
[46] E. Yuan, J. Tong, "Attributed based access control (ABAC) for web services" In IEEE International Conference on Web Services, 2005. https://doi.org/10.1109/ICWS.2005.25
[47] M. Bartoletti P. Degano, G. L. Ferrari, "Plans for service composition" in Workshop on Issues in the Theory of Security, 2006.
[48] S. Haibo, H. Fan, "A context-aware role-based access control model for web services" in IEEE International Conference on e-Business Engineering, 2005, pp. 220-223. https://doi.org/10.1109/ICEBE.2005.1
[49] S. M. Babamir, F. S. Babamir S. Karimi, "Design and evaluation of a broker for secure web service composition" in IEEE International Symposium on Computer Networks and Distributed Systems, 2011, 222-226. https://doi.org/10.1109/CNDS.2011.5764577
[50] M. Anisetti, C. A. Ardagna, M. Bezzi, E. Damiani, S. P. Kaluvuri, A. Sabetta, "A Certification-Aware Service-Oriented Architecture" Bouguettaya A., Sheng Q., Daniel F. (eds) Advanced Web Services. Springer, New York, NY, 2014, 147-170.
[51] J. Biskup, B. Carminati, E. Ferrari, F. Muller, S. Wortmann, "Towards secure execution orders for composite web services" in IEEE International Conference on Web Services, 2007, 489-496.
[52] B. Zhou, D. Llewellyn-Jones, Q. Shi, M. Asim, M. Merabti, D. Lamb, "A Compose Language-Based Framework for Secure Service Composition" in IEEE International Conference on Cyber Security, 2012, 195-202. https://doi.org/10.1109/CyberSecurity.2012.32
[53] F. Martinelli, I. Matteucci, M. Petrocchi, L. Wiegand, "A formal support for collaborative data sharing", in International Conference on Availability, Reliability, and Security, Springer Berlin Heidelberg, 2012, 547-561.
[54] C. K. Yee, Design and Implementation of Test Case Generation Tool for BPEL Unit Testing (Doctoral dissertation, thesis), 2007.
[55] H. Roth, J. Schiefer, A. Schatten, "Probing and monitoring of WSBPEL processes with web services" in 3rd IEEE International Conference on E-Commerce Technology, 8th IEEE International Conference on and Enterprise Computing, E-Commerce, and E-Services, 2006, 30-30. https://doi.org/10.1109/CEC-EEE.2006.69
[56] L. Baresi, C. Ghezzi, S. Guinea, "Smart monitors for composed services" in ACM 2nd international conference on Service oriented computing, 2004, 193-202. https://doi.org/10.1145/1035167.1035195
[57] Active Endpoint, ActiveBPEL. www.active-endpoints.com/active-bpel-engineoverview, 2007. htm, last accessed 2019/11/04
[58] L. Jamen, SGhosh, "Oracle Fusion Middleware Performance and Tuning Guide 11g" https://docs.oracle.com/cd/E21764_01/core.1111/e10108/bpel.htm#ASPER99175, last accessed 2019/11/04
[59] C. Johnson, B. Newport, "Develop high performance J2EE threads with WebSphere Application Server" http://www.ibm.com/developerworks/websphere/techjournal/0606_johnson/0606_johnson.html, last accessed 2019/11/04
[60] M. Oriol Hilari, J. Marco Gómez, J. Franch Gutiérrez, D. Ameller, "Monitoring adaptable SOA systems using SALMon" in 1st Workshop on Monitoring, Adaptation and Beyond, 2008, 19-28.
[61] F. Barbon, P. Traverso, M. Pistore, M. Trainotti, "Run-time monitoring of instances and classes of web service compositions" in IEEE International Conference on Web Services, 2006, 63-71. https://doi.org/10.1109/ICWS.2006.113
[62] D. Manova, I. Manova, S. Ilieva, D. Petrova-Antonova, "faultInjector: A Tool for Injection of Faults in Synchronous WS-BPEL processes" in IEEE 2nd Eastern European Regional Conf. on the Engineering of Computer Based Systems, 2011, 99-105. https://doi.org/10.1109/ECBS-EERC.2011.23
[63] D. Petrova-Antonova, D. Manova, S. Ilieva, "TASSA: Testing Framework for web service orchestration". in IEEE/ACM 10th International Workshop on Automation of Software Test, 2015, 8-12. https://doi.org/10.1109/AST.2015.9
[64] D. Petrova-Antonova, S. Ilieva, D. Manova, "Automated Web Service Composition Testing as a Service" in Hammoudi S., Pires L., Selic B., Desfray P. (eds) Model-Driven Engineering and Software Development. MODELSWARD 2016. Communications in Computer and Information Science, vol 692. Springer, Cham, 114-131.
[65] I. Spassov, D. Petrova, V. Pavlov, S. Ilieva, "Data Dependency Analysis Tool for Web Service Business Processes" Murgante B., Gervasi O., Iglesias A., Taniar D., Apduhan B.O. (eds) Computational Science and Its Applications - ICCSA 2011. ICCSA 2011. Lecture Notes in Computer Science, vol 6786. Springer, Berlin, Heidelberg, 232-243.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 169-173 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

A Study on Development of Information and Communication Ethics Sensitivity Measurement for Elementary School Students

Woochun Jun*

Seoul National University of Education, Department of Computer Education, 06639, Korea

ARTICLE INFO

Article history:
Received: 17 November, 2019
Accepted: 22 December, 2019
Online: 22 January, 2020

Keywords:
ICT Ethics
Sensitivity
Elementary School Students
ICT Ethics
Software Education

ABSTRACT

With advances in information and communication technologies and internet technologies, we have many benefits in our daily life. Most people have smart devices and use them for their work and leisure, and so on. However, these advances technologies brought us various side effects such as copyright violation, internet addiction, cybercrimes, personal information infringement, etc. In order to cope with those side effects, it is very important to teach information and communication ethics to students as early as possible. In addition, in order to teach young students, first, their awareness to information and communication ethics needs to be tested. In the literature, most awareness test tools are to test knowledge of information and communication ethics. In fact, those awareness tools have not consider sensitivity for information and communication ethics. In this work, an extensive and integrated sensitivity measurement tool is developed. At first sensitivity measurement standards are developed based on literature review works. Many elementary school teachers and ICT education experts survey the initial standards. Finally, sensitivity measurement standards are finalized with 6 areas and 19 standards. It is hoped that the proposed standards will be very helpful to diagnose awareness of elementary school students for information and communication ethics.

1. Introduction

In current knowledge and information society, everybody can have benefits from advances in information and communication technologies(ICT), internet technologies, and smart technologies, etc. Those advances technologies have brought many benefits in our daily life. Now many people have their own personal and smart device such as tablet PC, and can access to news and stock prices as quickly as possible.

Education area is also changing very quickly. In current classrooms, many PC and mobile devices are used for students and teachers. Students can get information anytime anywhere and exchange their ideas with anybody in the world. Recently, with advanced VR(Virtual Reality) and AR(Augmented Reality) technologies, VR/AR-based education becomes popular[1,2,3].

As we have more benefits from advanced technologies, side effects form those technologies become serious. The typical side effects include personal information infringement, copyright violation, cybercrimes, hacking, and so on. Most countries provide precaution and cure plans for those who suffering side effects. Usually costs for cure are tremendous. Thus, it is necessary to prevent those side effects as early as possible. The fundamental step to prevent is through education. Many countries provide education in the form of information and communication ethics education.

Recently robots are used in many areas as artificial intelligence and machine learning technologies have been developed. In this sense, ICT ethics education is extended to robot ethics education[4,5,6]. Robot ethics education can have big interest in the future as robots will be used more in industry and home. We need more fine and detailed principles and codes of ethics for more advanced robot.

In order to prevent and cure side effects from advanced technologies, it is very important to diagnose the status and awareness of students for ICT ethics. Based on correct diagnosis, it is possible to provide appropriate plans for precaution and cure. In the literature, awareness measurements are based on sole knowledge level of students. Knowledge on ICT ethics is important, however, sensitivity is more important. Sensitivity is defined formally in the next section.

*Corresponding Author: Woochun Jun, Seocho Dong, Seocho Gu, Seoul, Korea, wocjun@snue.ac.kr

The purpose of this work is to develop information and communication ethics sensitivity measurement standards for elementary school students. For this purpose, first, the initial sensitivity measurement standards are developed. Then, the proposed standards are surveyed by 30 elementary school teachers majoring information education. Finally, 19 sensitivity measurement standards with 6 areas are developed.

The rest of this paper is organized as follows. In Chapter 2, related works are introduced. In Chapter3, information and communication ethics sensitivity measurement standards for elementary school students are developed. In Chapter 4, statistical analysis is presented for those measurement standards. Finally, in Chapter 5, conclusions and further research issues are presented.

2. Related Works

2.1. Definitions of Sensitivity

Sensitivity is not formally defined in ICT ethics education. However, sensitivity is used for other ethics such as nursing ethics education. In nursing ethics education, moral sensitivity is defined as follows: Moral sensitivity is the ability to sensitize and detect conflict by sensitively identifying the problem surrounding the patient [7,8].

In this paper, sensitivity in information and communication ethics is defined as follows: Sensitivity is the ability to judge responsibilities for information communication ethics sensitively and recognize the consequences of their actions

Sensitivity is more important than knowledge on ICT ethics since knowledge does not relate with actions. That is, someone knows danger of copyright violation, however, he or she can keep doing misconducts.

2.2. Literature Review

In the literature, there are some works on awareness measurement standards of ICT ethics for students[9,10,11,12,13]. However, these works are concerned with awareness on knowledge. It means that the awareness measurement standards do not reflect sensitivity.

In this paper, some of awareness standards are introduced as follows.

In [11], information and communication awareness measure standards are proposed. The standards are developed based on TPB(Theory of Planned Behavior) that is widely used to explain the major variables affecting a person's behavior. The standards consist of 4 areas(moderation, respect, responsibility, and participation) and 57 detailed standards.

In [12], a measurement tool is developed for test consciousness of information and communication ethics for middle school students. The tool consists of 5 areas, respect, justice, etiquette, freedom, and responsibility, respectively. The tool also has final 33 standards after statistical analysis. The following Table 1 shows areas and description for each area in their work.

In [13], ICEI (Information and Communication Ethics Index) is proposed. They insist that, if the purpose of information and communication ethics education is to prevent and remedy the behavior of the information and communication ethics so that the youth themselves can judge the ethical and unethical in the information service society and do the right thing, then the development of the information and communication ethics index, which can measure the level of behavior of the information communication ethics problem of the youths, is more important than anything else. To ensure the reliability and feasibility of index developed in this study, it was applied to middle and high school students for an empirical analysis.

Table 1. Areas and Description of the Measurement Tool[12]

Area	Description
Respect	It is a tribute to the noble values of people and things in the information and communication society. This evaluates personal information protection and the perception of oneself and others as individuals.
Justice	This is a fair and correct way of doing things in the information and communication society. It evaluates the extent to which rules and laws are allowed to act properly and fairly.
Etiquette	This is the language etiquette and greeting etiquette that should be observed in the information and communication society. This evaluates the extent to which communication etiquette is observed and others are sought.
Freedom	This is a self-regulating and proactive attitude that does not harm others in the information and communication society. This evaluates the degree of self-control and self-regulation of one's will.
Responsibility	This is, in information and communication society, responsibility for its actions. It evaluates the extent of the act of disseminating information and the individual's own rules.

3. Development of Evaluation Metrics for Learners

3.1. Design Principles

The sensitivity measurement standards are developed with the following principles.

First, in order to develop comprehensive and integrated standards, overall aspects of education elements in ICT ethics education are considered and reflected.

Second, sensitivity is emphasized rather than knowledge and action for developing standards in this work. Sensitivity reflect interest and will to stop unfair action for elementary school students.

Third, too detailed standards such as standards depending on the current popular smart devices are not considered. This is

because too detailed standards will be changed or removed depending on popularity of the device or technology.

3.2. The Original Sensitivity Measurement Standards

The initial standards are developed with 6 areas and 19 standards. The following Table 2 shows the original standards.

Table 2. The Original Selection Standards

Area	Standards
General	1. I am interested in information and communication ethics. 2. I am willing to practice the recommended acts of ethics in everyday life. 3. I will stop immediately if I know of any harmful behavior in ICT.
Netiquette	4. I'm interested in netiquette 5. I am willing to practice the netiquette's recommended behavior in everyday life. 6. I will stop immediately if I know of any harmful behavior on netiquette.
Personal Information Protection	7. I am interested in privacy. 8. I am willing to practice the recommended acts for the personal information protection in my daily life. 9. I will stop immediately if I know of my personal information violation.
Copyright Protection	10. I am interested in copyright protection 11. I am willing to practice the recommended acts for the copyright protection in my daily life. 12. I will stop immediately if I know of my copyright violation
Internet Addiction	13. I am interested in internet addiction. 14. I am willing to practice the recommended actions to prevent internet addiction in my daily life. 15. I will stop immediately if I know of an act or habit that could lead to internet addiction. 16. I will be treated and consulted if it turns out to be an internet addiction.
Cyber Crime	17. I am interested in cybercrime. 18. I have a willingness to practice free behavior from cybercrime in my daily life. 19. I will stop immediately if I know my behavior is a cybercrime.

4. Statistical Analysis of the Proposed Standards

4.1. Overall procedure

In order to test the reliability of the proposed standards, extensive statistical works are done. First, 30 elementary school teachers in Seoul Metropolitan area in Korea are surveyed. Their majors are all information education. They are surveyed on May and June 2019. They are supposed to answer 5 scales for each standard as follows.

-Very likely(5 point)

-Slightly likely(4 point)

-Average(3 point)

-Slightly not likely(2 point)

-Very not likely(1 point)

4.2. Item Analysis

Table 3 shows the mean, standard deviation, skewness, and kurtosis of the 19 standards to measure students' sensitivity to information and communication ethics.

Table 3. Results of Item Analysis

No	M	SD	Skewness	Kurtosis
1	3.93	1.01	-0.50	-0.88
2	4.27	0.94	-1.11	0.28
3	4.37	0.96	-1.82	3.80
4	4.07	1.14	-1.18	0.59
5	4.37	0.93	-1.38	1.03
6	4.33	1.03	-1.56	2.21
7	4.27	0.91	-0.87	-0.42
8	4.17	0.91	-0.64	-0.79
9	4.27	1.05	-1.35	1.48
10	4.27	0.91	-0.87	-0.42
11	4.27	0.87	-0.91	-.0.08
12	4.30	0.99	-1.58	2.77
13	4.47	0.90	-1.72	2.18
14	4.47	0.82	-1.50	1.63
15	4.57	0.90	-2.67	8.11
16	4.20	0.92	-0.99	0.19
17	3.87	1.14	-0.63	-0.98
18	4.00	1.11	-0.80	-0.69
19	4.17	1.12	-1.31	1.02

Where M means mean and SD means standard deviation

As a result of the analysis, the mean ranged from 3.87 to 4.57 and the standard deviation ranged from 0.82 to 1.14. In addition, as a result of analyzing the skewness and kurtosis to examine the distribution of the items, the skewness range of each standard was -0.50□ -2.67 and the kurtosis range was -0.08□ 8.11. In general, if the absolute value of the skewness index is greater than 3.0, the skewness is severe, and if the absolute value of the kurtosis index is greater than 10, it can be determined that there is a problem in the normal distribution. Given these criteria, the descriptive statistics and item distribution of the items to measure students' evaluation of information and communication ethics sensitivity are good.

4.3. Reliability Analysis

In order to verify the reliability of the items measuring the sensitivity of information and communication ethics, the results

of analyzing the correspondence between the items using the Cronbach α value are shown in Table 4.

Table 4. Results of Reliability Analysis

Area	Item No.	Correlation Coefficient	Removal Factor
General	1	0.10	0.76
	2	0.57	-0.19
	3	0.37	0.33
Cronbach's α = 0.51			
Netiquette	4	0.44	0.84
	5	0.72	0.53
	6	0.62	0.62
Cronbach's α = 0.75			
Personal Information Protection	7	0.73	0.80
	8	0.78	0.75
	9	0.69	0.85
Cronbach's α = 0.86			
Copyright Protection	10	0.66	0.88
	11	0.88	0.69
	12	0.70	0.85
Cronbach's α = 0.86			
Internet Addiction	13	0.53	0.91
	14	0.89	0.78
	15	0.89	0.77
	16	0.66	0.87
Cronbach's α = 0.87			
Cyber Crime	17	0.62	0.91
	18	0.87	0.67
	19	0.73	0.81
Cronbach's α = 0.86			
(Overall) Cronbach's α = 0.94			

As shown in Table 4, Cronbach α showed 0.51 for general, netiquette 0.75, personal information protection 0.86, copyright protection 0.86, internet addiction 0.87, cyber crime 0.86, and overall information ethics sensitivity of 0.94, all above 0.50. The reliability is very good.

Table 5. Results of Correlation Analysis

Area	Correlation Value
General	0.796***(0.000)
Netiquette	0.859***(0.000)
Personal Information Protection	0.824***(0.000)
Copyright Protection	0.881***(0.000)
Internet Addiction	0.845***(0.000)
Cyber Crime	0.781***(0.000)

*** p<.001

4.4. Correlation Analysis

Table 5 shows the correlation between the information communication ethics sensitivity subdomain and the information communication ethics sensitivity.

Overall information and communication ethics (r = .796, p <.001), netiquette (r = .859, p <.001), personal information protection (r = .824, p <.001), copyright protection (r =. 881, p <.001), internet addiction (r = .845, p <.001), and cyber crimes (r = .781, p <.001) showed a statistically significant correlation with the overall information communication ethics sensitivity. Therefore, it can be seen that the overall information and communication ethics, which measure the sensitivity of general information and communication ethics, and netiquette, personal information protection, copyright protection, internet addiction and cybercrime are stable.

5. Conclusions and Further Research Issues

In the modern knowledge and information society, everyone is benefiting from advanced information techniques. However, various side effects also become a big problem. These side effects are time consuming and expensive to prevent and treat. Therefore, prevention through school education is very important.

At present, a variety of information and communication ethics education is provided to young adolescents. In addition, various measures have been developed to properly educate youth. Most scales, however, measure knowledge or will of them. Those scales do not reflect how they are sensible for information and communication ethics.

In this paper, the comprehensive measurement standards for sensitivity are developed. After the initial standards are developed, those standards are finalized through rigorous statistical analysis. Finally, 19 standards with 6 areas are proposed. These standards will be very helpful for measuring information and communication ethics sensitivity for elementary school students.

The immediate research issues are as follows. First, usability of these standards need to be tested by applying them to elementary school students. Second, it is necessary to develop extensive standards for middle school students and high school students, and adults. Recently, ethical issues for artificial intelligence are getting attention in the 4th industrial revolution era[14,15]. In the long run, the proposed sensitivity measurement standards need to be extended for artificial intelligence and big data issues.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

This work was supported by the 2019 Research Fund of Seoul National University of Education

References

[1] P. Song, H. Ryou, “A Study on VR Curriculum”, Journal of Korea Internet of Things Society, **2**(4), pp. 21-25, 2016. https://doi.org/10.20465/KIOTS.2016.2.4.021

[2] G. Yoo, S. Kim, “Effects of STEAM activities using VR.AR on creative problem solving ability and peer interactions of children”, The Journal of

Korea Open Association for Early Childhood Education, **24**(2), pp. 525-560, 2019, http://dx.doi.org/10.20437/KOAECE24-2-20

[3] Y. Chung, "Development of VR Fire-extinguishing Experience Education Contents Using UX Design Methodology", The Journal of the Korea Contents Association, **17**(3), pp. 222-230, 2017, http://dx.doi.org/10.5392/JKCA.2017.17.03.222

[4] J. Park, J. Han, "Development of Educational Contents in Robot Ethics for Elementary Students", Journal of Creative Information Culture, **4**(2), pp. 143-151, 2018.

[5] E. Yoo, M. Cho, "The Ethics of Robots and Humans in the Post-Human Age", The Journal of the Korea Contents Association", **18**(3), pp. 592-600, 2018, https://doi.org/10.5392/JKCA.2018.18.03.592

[6] S. Byun, H. Shin, J. Jeong, H. Kim, "A Study on Necessity of The Chart of Robot Ethics and its Contents", Journal of Ethics, **1**(112), pp. 295-319, 2017, http://dx.doi.org/10.15801/je.1.112.201703.295

[7] M. Yoo, K. Shon, "Effects of Nursing Ethics Education on Biomedical Ethics Awareness, Moral Sensitivity and Moral Judgment of Nursing Students", Journal of the Korean Bioethics, **12**(2), pp. 61-76, 2012.

[8] M. Park, D. Kjervik, J. Cradell, M. Oermann, "The Relationship of Ethics Education to Moral Sensitivity and Moral Reasoning Skills of Nursing Students", Nursing Ethics, **19**(4), pp. 568-580, 2012, https://doi.org/10.1177/0969733011433922

[9] J. Lee, K. Han, Y. Lee, S. Kim, "Development of an Index Model on the Information and Communication Ethics", Journal of the Korean Society of Computer Education, **10**(3), pp. 19-29, 2007.

[10] S. Cho, S. Kim, "Analysis of Factors to Measure Information Communication Ethics of Adolescents Using Delphi and AHP", Journal of the Korean Society of Computer Education, **11**(9), pp. 1-9, 2008.

[11] S. Cho, "The Development of Scale based on TPB to Measure the Information Communication Ethics Awareness", The Journal of Curriculum and Evaluation, **15**(1), pp. 291-320.

[12] I. Cho, J. Kim, "Measurement Tool Development for the Consciousness of Information and Communication Ethics", The Journal of Educational Information and Media, **15**(4), pp. 107-128, 2009.

[13] K. Kim, S. Kim, "Development and Validation of the Information and Communication Ethics Index of Adolescents", The Journal of Curriculum and Evaluation, **17**(3), pp. 191-221, 2014.

[14] V. Dignum, "Ethics in Artificial Intelligence: Introduction to the Special Issue", Ethics and Information Technology, **20**(1), pp. 1-3, 2018. https://doi.org/10.1007/s10676-018-9450-z

[15] A. Jobin, M. Lenca, and E. Vayena, "The Global Landscape of AI Ethics Guidelines", Nature Machine Intelligence, 1, pp. 389-399, 2019, https://doi.org/10.1038/s42256-019-0088-2

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 174-182 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Classification Model of Contact Center Customers Emails Using Machine Learning

Michael Wenceslaus Putong[*1], Suharjito[2]

[1]*Computer Science Department, Binus Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480,*

[2]*Computer Science Department, Binus Online Learning, Bina Nusantara University, Jakarta, Indonesia 11480*

ARTICLE INFO

Article history:
Received: 31 October, 2019
Accepted: 04 January, 2020
Online: 22 January, 2020

Keywords:
Email Classification
Machine Learning
Text Mining

ABSTRACT

E-mail is one of the media services used at the contact center. The challenge faced by e-mail services is how to handle e-mails that enter large quantities every day efficiently to provide fast and appropriate service to customers. The purpose of this study is to find which method has the best accuracy in classifying emails with four classes. The machine learning models compared in this study are Naive Bayes, SVM, and KNN. The data used in this study are primary data got from one of the contact centers. The NLP technique - Stop word removal, Stemming, and feature extraction using TF-IDF and Word2vec also applied to each algorithm to improve accuracy. The results of this study indicate that the SVM model with the Word2vec data feature produces the highest level of accuracy and the lowest level of accuracy produced by the Naive Bayes model using the TF-IDF data feature. The conclusion is that the classification using the word2vec data feature has a better level of accuracy than the classification using the TF-IDF data feature.

1. Introduction

Email is one of the tools used to communicate today. Email usage has substantially increased globally. In 2015, the number of emails sent and received, reach over 205 billion per day, and expected to grow around 3% every year, and reach over 246 billion at the end of 2019 [1]. Due to the strong increase of internet penetration, many customers use email to substitute for traditional communication methods such as letters or phone calls. As a result, the company receives every day numerous emails. Previous studies only classify e-mail with two categories, namely spam, and not spam, while in the contact centre the categories used to verify e-mail are four, namely, complaint, inquiry, transaction, and maintenance. With the huge volume of emails received by the contact centre every day, it will be very difficult to process these emails quickly. Hopefully, this research can find the classification model with the best accuracy that applies to be used to assist in processing e-mail at contact centre, especially in terms of categorization. At present, companies are outsourcing their internal email management to a dedicated call-centre environment. Handling e-mail efficiently is one of the main challenges in business [2]. This paper describes the methodologies method that can classify emails into four different categories based on the category that has applied in the contact centre that is, complaint, inquiry, maintenance, and transaction. The dataset used in this research is data primer collected from one of the contact centre. The dataset through the pre-processing stage before the accuracy, precision, and recall of each algorithm evaluated. Data cleaning, case folding, tokenizing, stemming and stop words elimination are pre-processing techniques that have widely used and combined with various algorithms to help improve and analyse which combinations give the best results [3]. The feature from documents extracted using TF-IDF. TF-IDF is a product of two statistics, namely Term Frequency and Inverse Document Frequency. To differentiate more, the number of terms that appear in each document calculated, and all added together [4].

2. Related Works

This paper focuses on comparing the algorithms to find the best result in classifying the emails based on the category used by the contact centre to classify customer emails. There are much research has been conducted for email classifying.

Harisinghaney proposed a research to detect spam emails based on text and images using three algorithms that is Naïve Bayes, KNN and Reverse DBSCAN. They adapt spam filters for each user's preferences and predict whether or not e-mails include

[*]Corresponding Author Michael W. Putong, Computer Science Department, Binus Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia, Email: michael.putong@binus.ac.id

spam using text mining and text recognizing with OCR library TESSERACT. in the study; they could achieve accuracy almost 50% better using pre-processed data compared to the accuracy achieved without using pre-processed data in all three algorithms. KNN with pre-processing data gets 83% accuracy in text and image-based spam filtering compared with 45% without pre-processing data. Similarly, Using Reverse DBSCAN, we achieved 74% accurate results using pre-processed data compared to 48% accuracy without pre-processed data. And finally, the best accuracy achieved by the Naive Bayes algorithm which is an 87% accurate result which is only 47% without pre-processing data [5].

Anitha used a Modified Naïve Bayes (MNB) algorithm to classify emails including spam or not spam. the results indicate that MNB is a spam email classifier that can classify with an average accuracy of 99.5%. Also, this requires a smaller amount of data for training and to provide standard performance with very low training time, 3.5 seconds. So far from this study, it was concluded that MNB is a fast and reliable classifier because it is related to the probability of words independent in the contents of an email. MNB provides the ethics of a new approach to email classification by combining probabilities independent of sequential words [6].

Gomes has studied a comparative approach to classify e-mails whether they are in the category of spam or non-spam e-mail using the Naïve Bayes Classifier and Hidden Markov Model (HMM). Categorization is done by only considering the text content of the body of the email. the results showed that HMM for classification provides better accuracy [1].

The anti-spam email system was implemented by Esmaeili in their research, they implemented an anti-spam system using the Naïve Bayes vs. method. PCA as a classifier, to classify spam and non-spam emails and use the feature selection method to increase the strength and speed of the classifier. The results of the study show that the Bayesian method with less miss classification had better precision compared to PCA, but PCA is a very fast method compared to the Bayesian. So, by increasing the number of training emails, and also using a good classifier such as SVM or ANN instead of the 1-NN method can increase the power of the PCA method [7].

In this study the authors will compare the results of the accuracy of the classification of three methods, namely Naïve Bayes Classification, K-NN and SVM. If in previous studies only classify emails in two classes, namely spam or non-spam, in this study email will be classified in 4 classes, namely *complaints, inquiries, maintenance and transactions* according to the category used by the banking contact center to classify customer emails.

If in the previous studies using data sources that mostly come from Enron Corpus, but in this study the data used are primary data from the database of one of the banking contact centers. Furthermore, if in previous studies only classify emails into two classes, namely spam and non-spam emails, but in this study, emails are classified into four classes according to the contents of the email namely maintenance, complaint, transaction and inquiry. In this study also uses and compares two different data feature extraction methods namely tf-idf and word2vec, where in previous studies most of them only used one method to extract data features.

3. Research Method

This research is motivated by the development of the company's service business to customers through contact centers which currently not only serve through telephone media but also through other media, one of which is via email and how contact centers are able to provide fast services to process customer emails where at This is to categorize the customer's email is still done manually by the contact center agent. The stages of the research carried out can be seen in Figure 1.

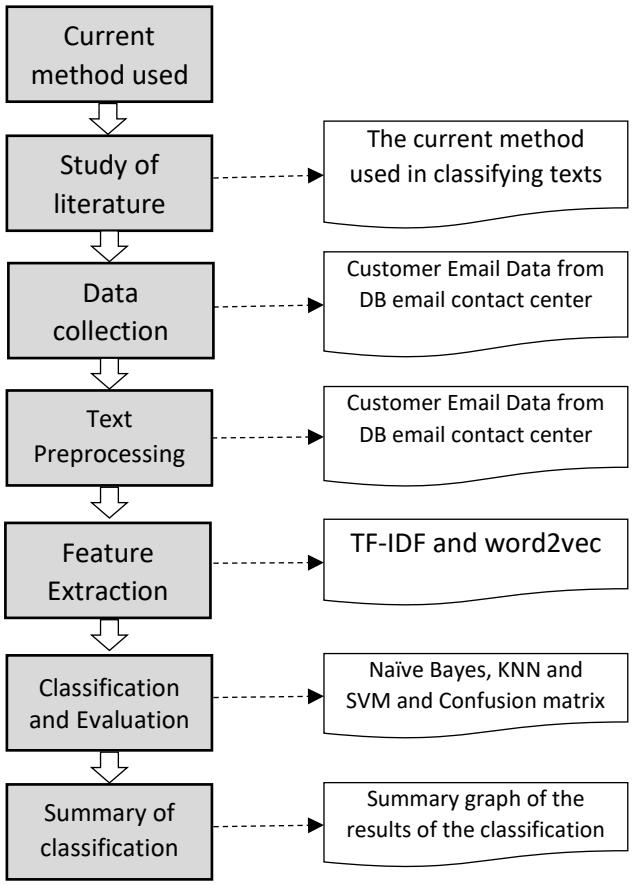


Figure 1: Research Stages

The data used in this study are primary data originating from the contact center email banking database, namely customer emails sent to the call center in the period 2016 to June 2018. The data is obtained by taking directly from the contact center email database.

3.1. Preprocessing

The data that has been obtained will go through the text preprocessing stage with the following methods [8] :

- Tokenization is the procedure of separating the text into words, phrases, or other important parts called tokens. In other words, tokenization is a form of text segmentation. Specifically, segmentation carries or considers only alphabetical or alphanumeric characters that separated from non-alphanumeric characters (for example, punctuation and spaces).
- Stop-words are words that commonly found in the text without dependence on certain topics (for example, conjunctions, prepositions, articles, etc.). Therefore, stop-words usually assumed to be irrelevant in the study of text classification and omitted before classification.

Specific stop-words for languages that are being studied, such as stemming.

- Convert into lowercase. At this step, it will convert all letters in the uppercase form into lowercase forms before classified.
- Stemming is to get the root word or the form of words that derived. Because words that semantically derived are similar to the root form, word events are usually calculated after applying stemming to the given text. Stemming algorithms are indeed specific to the language being studied.

3.2. Feature Extraction

Text classification is one of the main applications of machine learning. His job is to place new documents without labels into the specified categories. The text classification process involves two main problems, the first problem is the process of extracting feature terms that are effective in the training phase and the second is the actual classification of documents using feature terms in the test phase. Before classifying text, pre-processing has been done. In pre-processing Stop words are omitted and Stemmed is done.

Term frequency is calculated for each term in the document, and TF-IDF is also calculated [4].

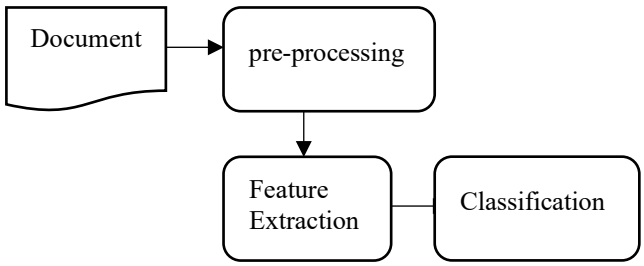


Figure 2: Document Classification Process with feature extraction

Term Frequency-Inverse Document Frequency (TF-IDF) is a numerical statistic that reveals how important a word is to a document. TF-IDF is often used as a weighting factor in information retrieval and text mining. The TF-IDF value increases proportionally to the number of times a word appears in a document but is contrary to the frequency of words in the corpus. This can help control the fact that some words are more common than others. TF-IDF can be successfully used to filter Stop-words in various subject areas including text summaries and classifications.

Term Frequency (TF) is defined as the number of times a term appears in a document.

$$\mathrm{tf(t,d)} = 0.5 + \frac{0.5 \text{ X f (t, d)}}{\text{Max num the occurrences of words}} \quad (1)$$

Inverse Document Frequency (IDF) is the statistical weight used to measure the importance of a term in a text document. The IDF feature is included where it reduces the weight of terms that often appear in the document and increases the weight of terms that rarely appear.

$$\mathrm{df(t,d)} = log + \frac{|\mathrm{D}|}{\text{num of doc where } terms\ (t) \text{ occured}}$$

Terms Frequency-Inverse Document Frequency (TF-IDF) is calculated using the following formula:

$$\mathrm{tfidf(t,d,D)} = \mathrm{tf(t,d)x\ idf(t,d)} \quad (3)$$

In word2vec, there are two main learning algorithms, continuous bag-of-words, and continuous skip-gram. With continuous bag-of-words, the sequence of words in history does not affect projections. This predicts the current word based on the context. Skip-gram predicts the surrounding words given by the current word. Unlike the standard bag-of-words model, continuous bag-of-words use distributed representation from the context. It is also important to state that the matrix of weights between the input and the projection layer is shared for all word positions. The skip-gram model by default has a training complexity architecture as follows:

$$Q = C\ x\ (D + D\ x\ \log_2(V)) \quad (4)$$

From the formula can be explained, C is the maximum distance for words, D is a representation of the word, and V is dimensionality. This means that for each training word, we will randomly select a number of R in the range <I; C> and use the word R from history and the word R from the future of the word chosen as the correct label. This requires us to do two classifications of the word R with the word chosen as input and each word R + R as the output. Using a binary tree representation of VOCAB the number of output units that require evaluation can go down to around log2 (V) [9].

3.3. Text Classification Techniques

In general, the text classification technique can be divided into two, The Statistical and Machine Learning approaches. Pure Statistical Techniques meet the hypotheses that are manually proclaimed, therefore, the need for algorithms is only minimal. Whereas Machine Learning techniques are specifically made for automation [10].

Naïve Bayes (NB), is a Bayes theorem oriented learning model that is very useful for learning tasks involving high dimensions of data, such as text classification & web mining. In general Bayesian models, classification is obtained by using dependencies (or conditional dependencies) between random variables. This process is usually time-consuming because examining the relationship between all random variables is a combinatorial optimization task. Alternatively, Naïve Bayes loosens the structure dependence between attributes by simply assuming that the attributes are conditionally independent, given a class label. As a result, examining the relationship between attributes no longer needed and derivatives of the NB model can be linearly scaled to training data [11].

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\ P(A)}{\mathrm{P(B)}} \quad (5)$$

K-Nearest Neighbours (KNN) is an example-based classification algorithm where documents that are not seen are classified with the majority category k the most similar training documents. The similarity between two documents can be

measured by Euclidean distance from n feature vectors representing documents [12].

$$\text{Euclidean} \sqrt{\sum_{i=1}^{k}(X_i - Y_i)^2} \tag{6}$$

Support vector machine (SVM) is a class of machine learning algorithms that can do pattern recognition and regression based on statistical learning theory and the principle of structural risk minimization. Vladimir Vapnik created the SVM to look for a hyperplane that separates a set of positive examples from a set of negative examples with maximum margins. Margin defined by the distance from the hyperplane to the closest positive and negative examples [13].

$$\frac{1}{2}\ w^T\ w + C\sum_{i=1}^{N}\xi_i \tag{7}$$

3.4. Classification and Evaluation

The data ratio is used 80% for training data and 20% for testing data. In this stage the text classification will be carried out using the Naïve Bayes method, k-NN and SVM and comparing the accuracy values from the classification results of each method to determine which method has the best accuracy. Classification is divided into 4 classes according to categories namely, Complaint, Maintenance, Inquiry and Transaction.

The results of the text classification process will be evaluated to determine the accuracy of each classification method used. The classification results are displayed in the accuracy and confusion matrix table.

The formula for calculating accuracy, precision, recall and F1-score in a multi-class classification is as follows:

$$Accuration\ = \frac{\sum_{i=1}^{l}\frac{TP_i\ +\ TN_i}{TP_i\ +\ TN_i\ +\ FP_i\ +\ FN_i}}{l}\ x\ 100\% \tag{8}$$

$$Precision\ = \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l}(FP_i\ +\ TP_i)}\ x\ 100\% \tag{9}$$

$$Recall\ = \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l}(FN_i\ +\ TP_i)}\ x\ 100\% \tag{10}$$

$$F1 - Score\ = \frac{2\ x\ (Precision\ x\ Recall)}{Precision + Recall} \tag{11}$$

Where, TP_i is True Positive, TN_i is True Negative, FP_i is False Positive, FN_i is False Negative and l is the number of class classified.

A summary of the classification results will display a graph showing the comparison of accuracy, recall, precision and f1-score of the classification results for each model used in this study.

4. Result and Analysis

This research uses primary data originating from a banking contact centre that contains 55281 emails with different amounts of data for each label according to the amount of data got within the 2016 to 2018 period. The email data used has been manually labelled by contact centre agents based on the categories that have been determined by regulations that apply to the contact centre. Email is divided into 4 classes, namely, Maintenance, Inquiry, Complaint, and Transaction. Emails are labelled based on the intent and purpose contained in the body contents of the email. The following is an example of the email data used in this research.

Data split into training and testing data with ratio 80% for training and 20% for testing.

4.1. Pre-Processing

The following are the steps taken in pre-processing email data :

1. *Lowercase Conversion*
   At this step, all letters in the email transformed into lowercase letters.
2. *Stemming*
   In this step, each sentence in the body of the email is separated into words, according to the words that make up the sentence. The stemming process is done using the literary library in python.
3. *Tokenization*
   At this step, each sentence in the body contents of the e-mail is separated into words, according to the words that form the sentence.
4. *Remove Stop words*
   At this step, we eliminate all words that are not important or do not affect the data class.

4.2. Feature Extraction

The feature extraction process using the TF-IDF method produces 665 word features. Examples of feature extraction results using the TF-IDF method can be seen in Table 1.

Table 1: Sample of Feature Extraction Data Result Using TF-IDF

No	Word	Total Occurrences	Document Occurrences
1	adu	12,67	9,29
2	agenda	0,17	0,04
3	akibat	4,04	3,63
4	akses	2,33	1,96
5	akta	1,75	1,54
6	akte	0,29	0,25
7	aktif	29,17	19,29
8	aktifkan	0,04	0,04
9	aktivasi	5,46	3,67
10	akumulasi	0,38	0,38

The feature extraction process using the word2vec method is done with the parameters min_vocab_frequency = 10, and layer_size = 50. The min_vocab_frequency parameter is the minimum frequency of the number of words present in a document and layer_size is the number of vectors generated. The model will ignore words that do not meet the minimum number. The feature used is the average value of each word vector element

The result of feature extraction using word2cev produces 100 word features. An example of the feature extraction using the word2vec method can be seen in Table 2.

Table 2: Sample of Feature Extraction Data Result Using Word2vec

No	words	vector
1	kartu	-0,00029
2	kredit	-0,00408
3	mohon	0,00951
4	informasi	0,00858
5	kirim	0,02439
6	tagih	-0,02724
7	percaya	0,01411
8	hormat	0,00182
9	ucap	0,01586
10	surat	0,03402

4.3. Classification

The data classification in this study uses 10000 email data got from a database of one of the contact centers. Data is shared using split validation with a ratio of 80% for training data and 20% for testing data. The type of sampling used is stratified sampling. Email data consists of 4 classes that have 2500 emails for each class, namely Maintenance, Inquiry, Transaction, and Complaint. The data feature was extracted using the TF-IDF and word2vec methods.

A. Naive Bayes

Table 3 is the confusion matrix of the email classification results using the Naïve Bayes model and data feature extraction using the TF-IDF method.

Table 3: Confusion Matrix Naive Bayes model with TF-IDF feature extraction

	true complaint	true inquiry	true maintenance	true transaction	class precision
pred. complaint	146	37	21	0	71.57%
pred. inquiry	131	139	57	0	42.51%
pred. mainenance	162	246	230	0	36.05%
pred. transaction	61	78	192	500	60.17%
class recall	81.60%	34.20%	33.40%	100%	
Total Email	500	500	500	500	

From table 3 it can be explained that out of the total 2000 emails classified by the number of each class of 500 emails, 146 emails were predicted as true email complaints and 204 emails were predicted as false email complaints, 71.75% class precision and class recall 81.60%. There were 139 emails predicted to be true email inquiry and a total of 188 emails predicted to be the false email inquiry, class precision 42.51% and class recall 34.20%. 230 emails were predicted as true email maintenance and a total of 408 emails were predicted as false email maintenance, class precision 36.05% and class recall 33.40%. 500 emails were predicted as true email transactions and a total of 331 emails were predicted as false email transactions, 60.17% precision classes and 100% class recall.

Table 4 is the confusion matrix of the email classification results using the Naïve Bayes model and data feature extraction using the word2vec method.

Table 4: Confusion Matrix Naive Bayes model with Word2vec feature extraction

	true complaint	true inquiry	true maintenance	true transaction	class precision
pred. complaint	408	25	7	0	92.73%
pred. inquiry	18	171	64	0	67.59%
pred. maintenance	21	137	167	0	51.38%
pred. transaction	53	167	262	500	50.92%
class recall	81.60%	34.20%	33.40%	100%	
Total Email	500	500	500	500	

From table 4 it can be explained that out of the total 2000 emails classified by the number of each class of 500 emails, 408 emails were predicted as true email complaints and a total of 440 emails that were predicted as false email complaints, 92.73% class precision and class recall 81.60%. There were 171 emails predicted as true email inquiry and 82 emails predicted as false email inquiry, class precision 67.59% and class recall 34.20%. 167 emails were predicted as true email maintenance and a total of 158 emails were predicted as false email maintenance, class precision 51.38% and class recall 33.40%. 500 emails were predicted to be true email transactions and a total of 482 emails that are predicted to be false email transactions, class precision 50.92% and class recall 100.00%.

Table 5 and Figure 3 are tables and comparison diagrams of email classification results using the Naïve Bayes model and the TF-IDF and word2vec feature extraction method.

Table 5: Summary of Naive Bayes classification result

	Accuracy	Mean Precision	Mean Recall	F1-Score
TF-IDF	50,75%	52,57%	50,75%	51,65%
Word2vec	62,30%	65,65%	62,30%	63,93%

From table 8 and figure 2 above it can be seen that the accuracy of email classification using the Naive Bayes model combined with the word2vec feature extraction method has a higher accuracy rate of 63.30%, compared to the accuracy of the classification results of the Naive Bayes model combined with the TF-IDF feature extraction method. which is 50.75%.

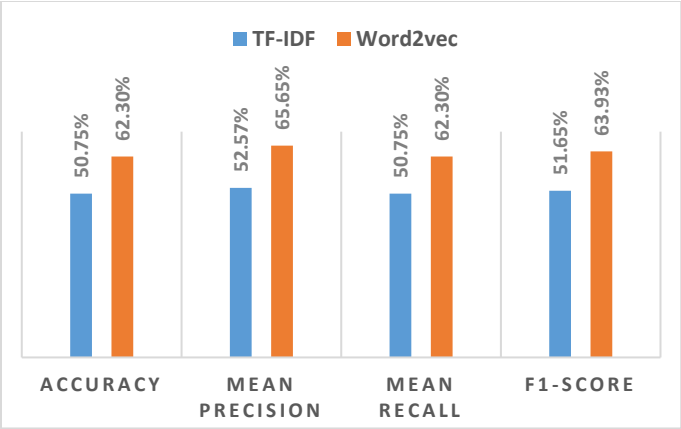


Figure 3: Summary of Naive Bayes classification result diagram

B. KNN

The K value used in this classification model is determined by testing using a different K value from the value of K = 1 to the value of K = 10. Figure 4.6 and Table 9 are diagrams and tables of the level of accuracy obtained from the test results with different K values. Classification is done by testing different measure types parameters. The highest accuracy results are obtained with parameters, Measures Types: Numerical Measures and Numerical Measures Type: Cosine Similarity.

Table 6: Level of Accuracy KNN Classification for each K value

k value	TF-IDF	word2vec
1	69,25	72,95
2	69,25	72,95
3	69,55	72,85
4	70,65	73,85
5	70	72,9
6	70,4	73,6
7	69,75	72,9
8	70,2	73,6
9	69,75	74,6
10	69,6	74,2

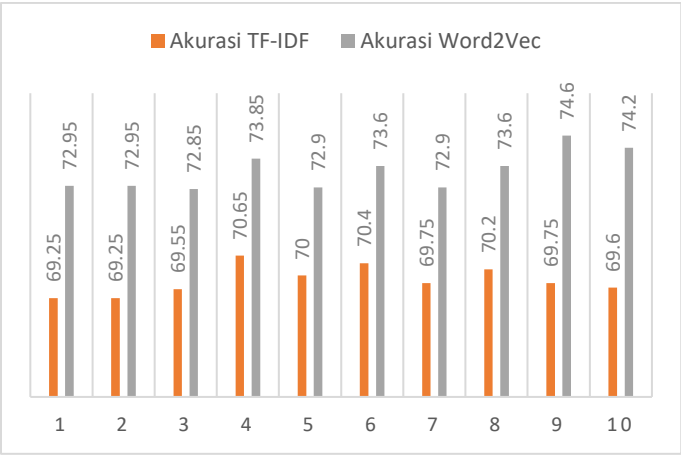


Figure 4: Level of Accuracy KNN Classification for each K value diagram

Table 7 is the confusion matrix of the email classification results using the KNN model with a value of K = 4 and data feature extraction using the TF-IDF method.

Table 7: Confusion Matrix KNN model with TF-IDF feature extraction

	true complaint	true inquiry	true maintenance	true transaction	class precision
pred. complaint	329	93	69	0	67.01%
pred. inquiry	108	290	134	0	54.51%
pred. maintenance	60	107	294	0	63.77%
pred. transaction	3	10	3	500	96.90%
class recall	65.80%	58.00 %	58.80%	100%	
Total Email	500	500	500	500	

From table 7 it can be explained, out of the total 2000 emails classified by the number of each class of 500 emails, 329 emails were predicted as true email complaints and a total of 162 emails were predicted as false email complaints, 67.01% class precision and class recall 65.80%. There were 290 emails predicted as true email inquiry and a total of 242 emails predicted as false email inquiry, 54.51% precision class and 58.00% class recall. 294 emails were predicted as true email maintenance and a total of 167 emails that were predicted to be false email maintenance, 63.77% precision class, and 58.80% class recall. 500 emails were predicted to be true email transactions and a total of 16 emails that are predicted to be false email transactions, 96.90% class precision and 100.00% class recall.

Table 8 below is the confusion matrix of the results of email classification using the KNN model with a value of K = 9 and data feature extraction using the word2vec method.

Table 8: Confusion Matrix KNN model with Word2vec feature extraction

	true complaint	true inquiry	true maintenance	true transaction	class precision
pred. complaint	333	50	24	0	81.82%
pred. inquiry	97	299	109	0	59.21%
pred. maintenance	58	135	360	0	65.10%
pred. transaction	12	16	7	500	93.46%
class recall	66.60 %	59.80%	72.00%	100%	
Total Email	500	500	500	500	

From table 8 it can be explained out of the total 2000 emails classified by the number of each class of 500 emails, 333 emails were predicted as true email complaints and a total of 74 emails were predicted as false email complaints, 81.82% class precision and class recall 66.60%. There were 299 emails predicted as true email inquiry and 206 emails predicted as false email inquiry, class precision 59.51% and class recall 59.80%. There are 360 emails predicted as true email maintenance and a total of 193 emails predicted as false email maintenance, 65.10% precision class and 72.00% class recall. 500 emails were predicted as true email transactions and a total of 35 emails were predicted as false email transactions, 93.46% class precision and 100.00% class recall.

Table 9 and Figure 5 are tables and comparison diagrams of email classification using the KNN model and the TF-IDF and word2vec feature extraction method.

Table 9: Summary of KNN classification result

	Accuracy	Mean Precision	Mean Recall	F1-Score
TF-IDF	70,65%	70,55%	70,65%	70,60%
Word2vec	74,60%	74,90%	74,60%	74,75%

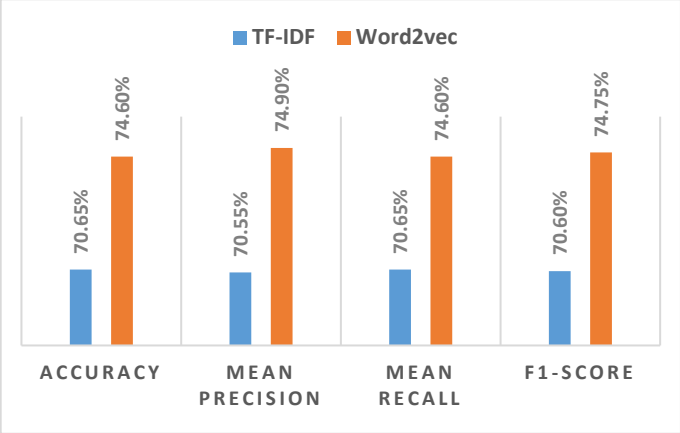


Figure 5: Summary of KNN classification result diagram

From table 9 and figure 5 above it can be seen that the accuracy of email classification using the KNN model using the word2vec data feature has a higher accuracy rate of 74.60% when compared to the KNN model using the TF-IDF data feature 70.65%.

C. SVM

Classification with the SVM model is done by testing different types of SVM. The highest accuracy is produced by the SVM model with C-SVC type, sigmoid kernel type and epsilon value of 0.001, which is 77, 85%. Table 13 is the configuration matrix of email classification results using the SVM model and data feature extraction using the TF-IDF method.

Table 10: Confusion Matrix SVM model with TF-IDF feature extraction

	true complaint	true inquiry	true maintenance	true transaction	class precision
pred. complaint	356	114	47	0	68.86%
pred. inquiry	107	305	163	15	51.69%
pred. maintenance	32	70	289	0	73.91%
pred. transaction	5	11	1	485	96.61%
class recall	71.20 %	61.00%	57.80%	97%	
Total Email	500	500	500	500	

From table 10 it can be explained out of the total 2000 emails classified by the number of each class of 500 emails, 356 emails were predicted as true email complaints and a total of 161 emails were predicted as false email complaints, 68.86% class precision and class recall 71.20%. There were 305 emails predicted as true email inquiry and 285 emails predicted as false email inquiry, class precision 51.69% and class recall 61.00%. 289 emails were predicted to be true email maintenance and a total of 102 emails that were predicted to be false email maintenance, 73.91% class precision and 57.80% class recall. 485 emails were predicted to be true email transactions and a total of 17 emails that were predicted to be false email transactions, class precision 96.61% and class recall 97.00%.

Table 11 is the configuration matrix of email classification results using the SVM model and data feature extraction using the word2vec method

Table 11: Confusion Matrix SVM model with Word2vec feature extraction

	true complaint	true inquiry	true maintenance	true transaction	class precision
pred. complaint	398	4	2	0	98.51%
pred. inquiry	56	311	114	0	64.66%
pred. maintenance	42	159	370	22	62.39%
pred. transaction	4	26	14	478	91.57%
class recall	79.60%	62.20 %	74.00%	95.60%	
Total Email	500	500	500	500	

From table 11 it can be explained out of the total 2000 emails classified by the number of each class of 500 e-mails, 398 e-mails were predicted as true e-mail complaints and a total of 6 e-mails were predicted as false e-mail complaints, class precision 98.51% and class recall 79.60%. There were 311 emails predicted as true email inquiry and a total of 170 emails predicted as false email inquiry, 64.66% class precision, and 62.20% class recall. 370 emails were predicted as true email maintenance, and a total of 223 emails were predicted as false email maintenance, 62.39% precision class and 74.00% class recall. 478 emails were predicted as true email transactions and a total of 44 emails were predicted as false email transactions, class precision 91.57% and class recall 95.60%.

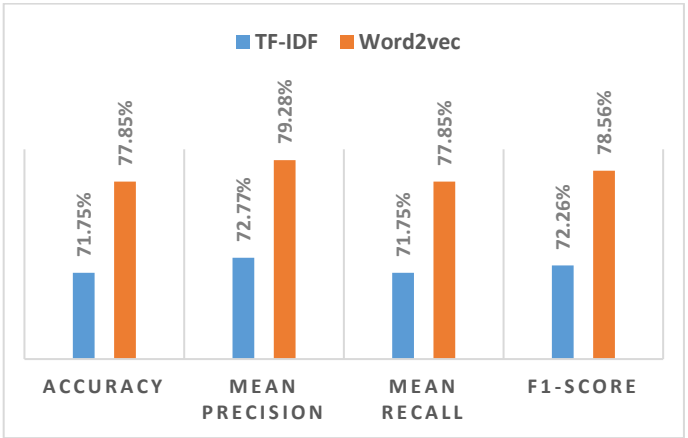


Figure 6: Summary of SVM classification result diagram

Table 12 and Figure 6 are a comparison of email classification results using the SVM model and data features obtained from the TF-IDF and word2vec methods. From table 12 and Figure 6 above it can be seen that the accuracy of email classification using the KNN model using the word2vec data feature has a higher accuracy

value of 77.85% when compared to the KNN model using the 71.75% TF-IDF data feature.

Table 12: Summary of SVM classification result

	Accuracy	Mean Precision	Mean Recall	F1-Score
TF-IDF	71,75%	72,77%	71,75%	72,26%
Word2vec	77,85%	79,28%	77,85%	78,56%

4.4. Classification Summary

Figure 7 shows the comparison of the accuracy value of the classification results of each model, the highest accuracy value generated by the SVM model with word2vec data features of 77.85%, and the lowest accuracy value generated by the Naive Bayes model with the TF-IDF data features of 50, 75%.

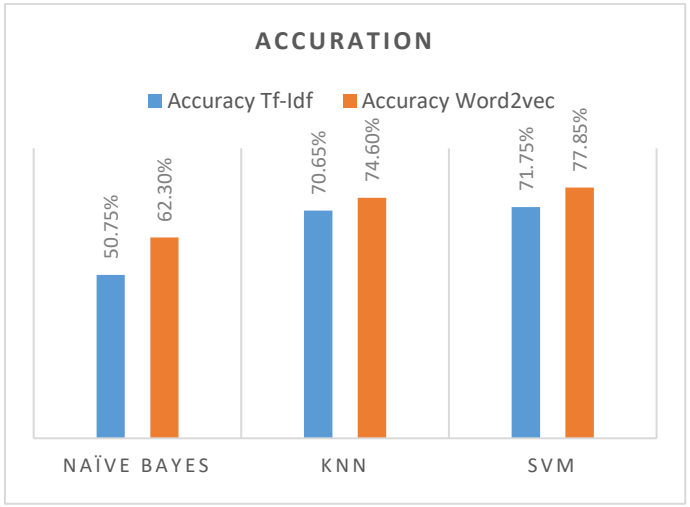


Figure 7: Comparison of Accuracy Diagram

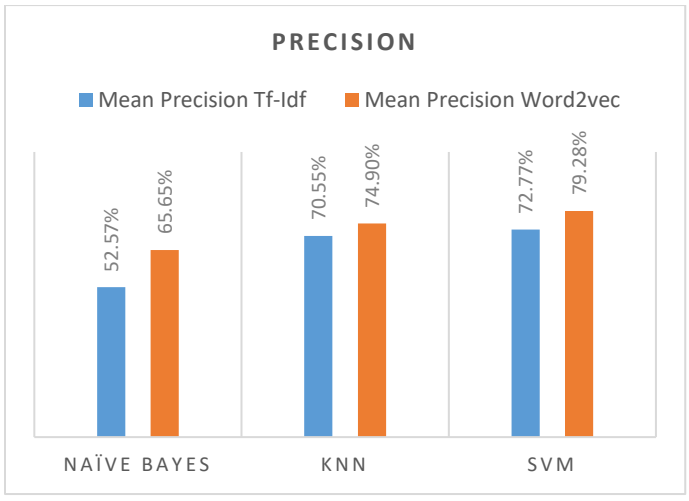


Figure 8: Comparison of Precision Diagram

Figure 8 shows the comparison of the average precision values from the results of the classification of each model, the highest average precision value generated by the SVM model with word2vec data features that is 79.28%, and the lowest average precision value produced by the Naive Bayes model with TF-IDF data features of 52.57%.

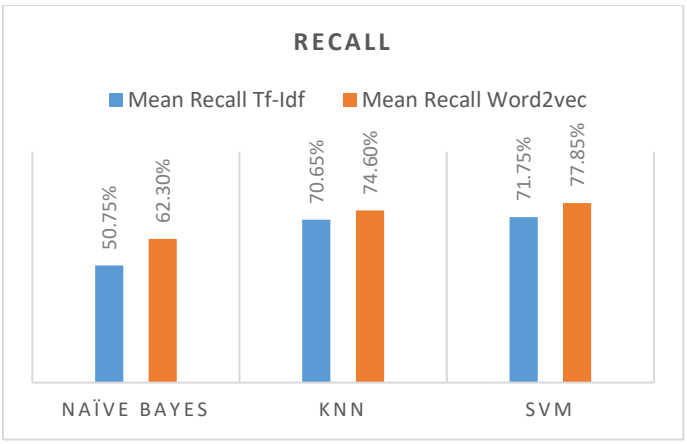


Figure 9: Comparison of Recall Diagram

Figure 9 shows a comparison of the average recall values from the classification results of each model, the highest average recall value generated by the SVM model with word2vec data features of 77.85%, and the lowest average recall value generated by the Naive Bayes model with TF-IDF data features of 50.75%.

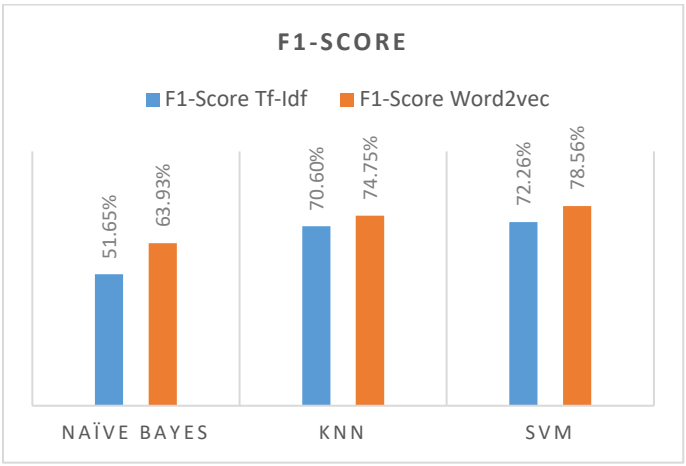


Figure 10: Comparison of F1-Score Diagram

Figure 10 shows a comparison of the F1-Score values from the classification results of each model, the highest F1-Score value generated by the SVM model with word2vec data features of 78.56%, and the lowest F1-Score value generated by the Naive Bayes model with the TF-IDF data features of 51.65%.

Overall accuracy values obtained by classification using the word2vec data features are better when compared to using the TF-IDF data feature. From the classification results, it can be concluded that the data features used in the classification affects the accuracy value.

5. Conclusion

Email classification using the SVM model with Word2vec data features has the highest accuracy rate of 77.85% and the lowest is Naive Bayes model using the TF-IDF data feature of 50.75%. From the results of the classification carried out by each model shows that, classification using different data features has an impact on accuracy, and classification using the word2vec data feature has a better level of accuracy than using the TF-IDF data feature.

References

[1] Gomes, S. R., Saroar, S. G., Mosfaiul, M., Telot, A., Khan, B. N., Chakrabarty, A., & Mostakim, M. (2017, September). A Comparative Approach to Email Classification Using Naive Bayes Classifier and Hidden Markov Model. In Advances in Electrical Engineering (ICAEE), 2017 4th International Conference on (pp. 482-487). Dhaka, Bangladesh: IEEE. doi:10.1109/ICAEE.2017.8255404

[2] Coussement, K., & Van den Poel, D. (2008). Improving customer complaint management by automatic email classification using linguistic style features as predictors. Decision Support Systems, 44(4), 870-882. doi:10.1016/j.dss.2007.10.010

[3] Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. International Journal of Computer Science & Communication Networks, 5(1), 7-16.

[4] Menaka, S., & Radha, N. (2013). Text classification using keyword extraction technique. International Journal of Advanced Research in Computer Science and Software Engineering, 3(12).

[5] Harisinghaney, A., Dixit, A., Gupta, S., & Arora, A. (2014, February). Text and image based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN algorithm. 153-155. doi:10.1109/ICROIT.2014.6798302

[6] Anitha, P. U., Rao, C. V., & Babu, S. (2017, November). Email Spam Classification using Neighbor Probability based Naïve Bayes Algorithm. In 2017 7th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 350-355). Nagpur, India: IEEE. doi:10.1109/CSNT.2017.8418565

[7] Esmaeili, M., Arjomandzadeh, A., Shams, R., & Zahedi, M. (2017, May). An Anti-Spam System using Naive Bayes Method and Feature Selection Methods. International Journal of Computer Applications, 165(4), 1-5.

[8] Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. Information Processing and Management, 50(1), 104-112. doi:10.1016/j.ipm.2013.08.006

[9] Lilleberg, J., Zhu, Y., & Zhang, Y. (2015, July). Support Vector Machines and Word2vec for Text Classification with Semantic Features. 136-140. doi:10.1109/ICCI-CC.2015.7259377

[10] Thangaraj, M., & Sivakami, M. (2018). Text Classification Techniques: A Literature Review. Interdisciplinary Journal of Information, Knowledge & Management, 13, 117-135. doi:10.28945/4066

[11] Wu, J., Pan, S., Zhu, X., Cai , Z., Zhang, P., & Zhang, C. (2015). Self-adaptive attribute weighting for Naive Bayes classification. Expert Systems with Applications, 42(3), 1487-1502. doi:10.1016/j.eswa.2014.09.019

[12] Adeva, J. G., Atxa, J. P., Carrillo, M. U., & Zengotitabengoa, E. A. (2014). Automatic text classification to support systematic reviews in medicine. Expert Systems with Applications, 41(4), 1498–1508. doi:10.1016/j.eswa.2013.08.047

[13] Zheng, B., Yoon, S. W., & Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Systems with Applications, 41(4), 1476-1482. doi:10.1016/j.eswa.2013.08.044

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 183-190 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Configurable Process Model: Discovery Approach from Event Logs

Rabab Sikal[*,1], Hanae Sbai[2], Laila Kjiri[1]

[1]*AlQualsadi research team, ENSIAS, Mohammed V University of Rabat, Morocco*

[2] *FST, University Hassan II of Casablanca, Mohammedia, Morocco*

ARTICLE INFO

Article history:
Received: 30 September, 2019
Accepted: 29 December, 2019
Online: 22 January, 2020

Keywords:
Configurable Process Model
Process Discovery
Variability

ABSTRACT

In the domain of business process management, the configurable process model is widely used to optimize time and cost of business process models design, which is known as the concept of "reuse". Using process mining techniques for process model discovery helps to provide a better view on processes and improve quality of models. The majority of existing configurable model discovery approaches work intensively on control flow discovery as main process perspective without considering other perspectives such as resources and data, and do not propose a detailed discovery of variability elements. In addition, the configurable process model creation is generally done by merging variant models not directly from event logs, which is not the optimal way to get a reliable configurable process model. This paper presents an overview of new multi-perspective variability discovery approach. The approach respects the variability of different process perspectives and allows users to create a configurable process model directly from event logs.

1. Introduction

Process models are designed, managed, configured mainly through methods and tools provided by the Business Process Management (BPM) domain [1,2]. Actually, organizations are increasingly opting for event systems, also called process Aware Information Systems (PAIS) for the purpose of analyzing, supervising and optimizing the organization's processes [3]. Nowadays, rapid increase of business needs and fast changing of the enterprise environment derive the enterprise to face the challenge of saving time, reducing costs and minimizing errors of process management. Therefore, opting for reuse concept is a big requirement for making an optimal and flexible business process design [4,5]. Hence, the importance of taking into account the previous design experiences and do not design processes from scratch. In this sense, different approaches have proposed the reuse concept in process design while flexibility and adaptability are addressed in business process models [6, 7].

The configurable process model is defined as a single model that assembles all process variants in one model. It is also called "customizable process model" which means that this kind of process model regroups options, which can be configured by users to derive desired process variants. It represents commonalities and differences between all process variants, which offers flexibility and enables process design through reuse concept. The variation point is a configurable element of configurable process model. It represents where the variation occurs in the process model and represents all possible design choices. The configuration of the model consists of making choices of options for each configurable element according to specific requirements in order to derive individual and suitable model for the enterprise with minimal design effort. However, despite the diversity of approaches proposing creation and configuration of configurable process models, their management still requires a significant manual work in different steps (e.g. design, configuration and evolution).

Against this background, the techniques of process mining are introduced with the aim to automate the process management and minimize human intervention. The process mining uses data recorded in the event logs during the process execution in order to help organizations for discovering, checking conformance and enhancing their business processes.

Creating process model manually is a hard and redundant task, since the use of similar processes becomes more and more popular. For that reason, configurable process model discovery is used as an alternative for reusable process design. In this context, several

[*]Corresponding Author Rabab Sikal, rababsikal@gmail.com

works have proposed approaches for configurable process discovery [8, 9, 10]. However, existing works discover the control-flow as the main process perspective without considering other process perspectives like data and resources. In addition, the proposed variability discovery approaches do not present an explicit and detailed discovery of variability elements, namely variation points and variants. Moreover, existing approaches for configurable model creation use algorithms to construct configurable model by merging similar variant models, which result in large and complex models. In the light of these limitations, we have proposed in [11,12] a multi-perspective configurable process discovery approach with respect to variability of activities and resources. The comparative studies presented in these two works showed the lack of support for variability discovery for various process perspectives. This paper extends the work originally presented in the 2018 IEEE 5th International Congress on Information Science and Technology (CiSt) [11] that we complete by proposing an approach for configurable process model creation directly from event logs. The organization of this paper is as follows: section 2 defines the background of our research field, while section 3 reviews approaches working on configurable process model discovery. In section 4, we present our proposed configurable process discovery approach. Section 5 depicts our multi-perspective discovery framework and presents the algorithms for variability extraction and configurable model creation. Finally, Section 6 presents conclusions and some future directions.

2. Background

This section introduces three basic concepts used in this paper: process discovery, configurable process model, and variability. Then, we present briefly the four configurable process model discovery approaches described in [8].

2.1. Process discovery

Process mining is a set of techniques applied to extract data recorded in event logs. These data concern all process actions captured during process execution and are used to discover, monitor, and improve processes [1]. There are three main areas of process mining [1]:

- Process Discovery: the discovery algorithm takes an event log in input and produces a process model in output without using any additional knowledge.
- Conformance checking: verifies if an existing or discovered process model fits to its event log, or vice versa.
- Enhancement: extends and enrich existing process model, already discovered, by using information recorded in event logs.

In our approach, we focus on process discovery. There are two main kinds of process discovery:

Process discovery from one event log: it is the classical type of discovery. It allows for extracting one process model for each event log. However, this kind generates redundant processes [13,14].

Process discovery of a collection of event logs: this concerns the discovery of configurable process model. It requires firstly regrouping all event log that can belong to the same family and then applying techniques to discover the configurable process [8,9].

In Our work, we are interested in the configurable process model discovery from a collection of event logs.

2.2. Configurable process model & variability

The configurable process model represents shared/non configurable and unshared/configurable parts by all process variants in one global model. The configuration of configurable parts depends on the needs and the various constraints specific to the organization [15, 16]. Indeed, modeling all process variants and updating common process items cause redundancies and errors. Hence, the choice of configurable processes, generally presented by the merge of multiple process variants into a single process, is very useful to facilitate reuse and manage variability [17, 18]. Different extensions of process modeling languages have been developed for configurable process models representation, namely C-BPMN, C-EPC [15], C-YAWL [16] and configurable process tree.

The configuration consists of deriving individual process models corresponding to the different process variants from the configurable model. This operation is called individualization. The individualization is about blocking a given path of the model, so it cannot be taken or hiding activities they it can be skipped during the process execution. Designing a configurable process model consists in defining all the different variants of a given business process first and then integrating all of them into a single configurable model.

The variability is a key concept for configurable process model creation. It is represented by two main elements, namely, variation point and variants [13]. The variability defines and manages variable elements of business process [17]. Therefore, we define a process model that supports organizational, behavioral, functional and informational variability, as a multi-perspective configurable process model [1]. This can make the configurable process more explicit and valuable.

2.3. Approaches for configurable process model discovery

Process mining techniques for automated process discovery use data recorded in the event logs to represent the process behavior through a process model [6, 1]. Indeed, applying mining techniques for configurable process discovery is very useful, given the time saved and the effort reduced compared to conventional methods. In the literature, different approaches of configurable process discovery have been proposed. Buijs et al. [8] proposed four configurable process model discovery approaches, presented as follow:

- **Approach 1:** it is an approach initially proposed by Gottshalk [14]. The configurable process model discovered with this approach is the result of merging process models discovered from each event log.
- **Approach 2:** with the aim to improve the approach 1, the approach 2 merges all event logs and uses them to discover a common process model. Then, for every event log an

individual model is generated. Finally, the construction of the configurable model is done by merging the individual models.

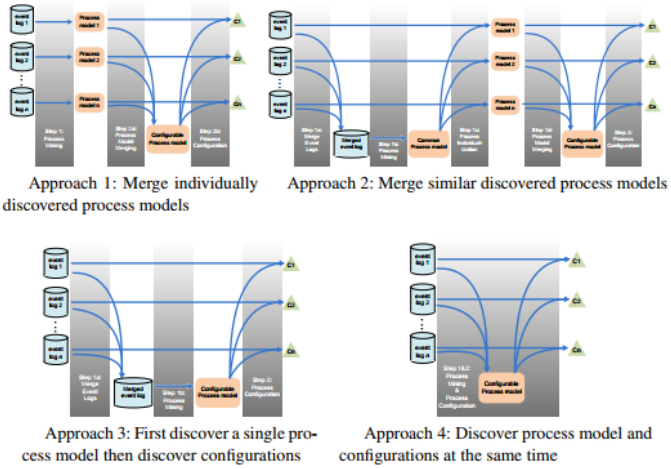


Figure 1: the four approaches of configurable process model discovery [8]

- **Approach 3:** it suggests the merging of different event logs into one merged event log. Then, the configurable is discovered. It captures the behavior of all model that describes the behavior of these event logs.
- **Approach 4:** it allows for the discovery of the process model and its configurations at the same time [8].

In this paper, our discovery approach belongs to the approach 4. It deals with redundancies and brings more flexibility in using discovery techniques to construct processes that capture variability.

3. Related works

In this section, we present several existing approaches for configurable process model discovery. Then, these approaches are evaluated according to four criteria. Finally, we discuss results and limitations.

3.1. Configurable process model discovery works

Several BPM studies are interested in using the paradigm of the "design by reuse" for the construction of a configurable process model. Some of them have proposed to construct the configurable process model by merging all process variants models into one model. Others have proposed to create a configurable process model using mining techniques on a collection of event logs. The author of [9] uses trace-clustering method for configurable process model discovery from collection of event logs. In [10], the author discovers configurable process fragments to avoid complex and large models. The author of [14] presents an approach using process mining and analysis techniques to merge two business process models into a single model for further process optimization. The approach in [16] merges the models of process variants to create configurable process model based on log files from various systems. The work provides suggestions for common and individual configurations. In [19], the author proposes two algorithms, one to compute merged models. The other, to extract digests from a merged model. The work [20] proposes an algorithm for constructing a configurable process by merging the process model of each variant, the process model generated by the algorithm is pre-annotated for the configuration step. The study of [21] splits the event logs in a cluster and for each cluster, A process model can be discovered. In case of large configurable process model, the model is reduced into a sub process model. Each sub-process model is configured independently to improve performance and to reduce complexity.

3.2. Comparative study

To summarize the previous section, the table 1 presents the principal points related to our approach and reached by every work. The approaches are evaluated according to four criteria defined as follows:

- **Variability discovery:** it indicates if the approach discovers explicitly the elements of variability (e.g. variation point, variants and variables).
- **Perspective discovery**: it presents the perspectives discovered by the approach. The main process perspectives are: control flow (C.F), resource (R), data (D) and configuration (C).
- **Configurable Model construction**: it indicates if the approach constructs configurable model.
- **Discovery approach for configurable model construction:** it indicates which approach, from the four approaches proposed by [8], is used for the construction of configurable model.

Table 1: comparative study

Works	Variability discovery	Perspective discovery				Configurable Process model construction	Discovery approach for configurable model construction			
		C.F	R	D	C		App 1	App 2	App 3	App 4
[9]		+	-	-	+	+	-	-	+	-
[10]	+	+	-	-	+	+	-	+	-	-
[14] [16] [19]	-	+	-	-	-	+	+	-	-	-
[20]	-	+	-	-	-	+	-	+	-	-
[21]	-	+	-	-	+	+	-	-	+	-

Table 1 is a summary of the evaluation criteria developed by each of the approaches presented in this section.

Variability discovery: The discovery of variability elements, namely variation point and variants, is present in few works. The works [10] discovers variability for configurable fragment of the process. The other studies [9, 14, 16, 19 20, 21] don't focus on variability elements in their discovery approaches.

Perspectives discovery: The studies [9, 10, 14, 16, 19, 20, 21] are limited to the discovery of control flow as a main process perspective, and discovery of its configurations. Thus, we notice the absence of support for discovery of other process perspective like resource and data.

Configurable process model construction: all the presented approaches propose the construction of configurable process model using different modeling language, like C-BPMN, C-EPC or process tree.

Discovery approach for configurable model construction: different works adopted different approaches for model construction. The works [14, 16, 19] construct configurable model based on approach 1. The [10, 20] adopt the approach 2 while [9,21] use approach 3.

3.3. Discussion

The analysis of the presented approaches shows that the majority of them do not present an explicit discovery of elements of variability and still limited to control flow discovery as a main perspective. While, other perspectives like resource still neglected and not integrated in the discovery approach. In addition, the construction of configurable process model is generally based on approach 1, 2 or 3 using merging techniques. We notice the absence of approaches for configurable model discovery based on the approach 4.

The limitations we conclude are as follow:

- The need for detailed discovery of variability elements, e.g. variation points, variation point types and variants. Whereas, discovering an explicit and detailed variability can be used to build configurable process models as well as its configurations. It can also be archived for potential process changes or improvements.
- Lack of multi-perspective discovery of process elements, namely data and resources. Instead of remaining focused on the analysis of the control flow, the extension of the configurable model with different process perspectives is of great importance. It helps analysts to manage the evolution of business process and to improve decision-making.
- The construction of the configurable model is based on merging individual models. Unlike, an approach that discovers configurable model directly from the event logs without merging individual models may provide a better model structure and better configuration options.

In the light of the presented limitations, we propose an approach with a detailed discovery of variability elements for different perspectives (control flow, resource). In order to enhance the variability discovery for other perspectives, we generate variability specification files for detailed variability. This can ensure traceability and optimize the process of changing or updating business process. In addition, our approach adopts the approach 4 proposed in [8] for optimal creation of configurable process model.

4. Preliminaries

In this section, we present formal definitions of basic concepts related to this work.

4.1. Event logs

Event logs are defined as files that store process data collected during process execution. The process mining techniques use data recorded in event logs for discovering process models, checking conformance between process model and its event log, detecting execution deviations or errors and observing social behaviors.

The table 2 illustrates an example of event log for a "purchase online" process. To buy an article the customer starts with "creating a personal account online (a)". After, the customer "choose products to buy (b)" and then "chose the payment method (c)" it can be "payment by card (d)", "payment by PayPal (e) " or "bitcoin payment (f)". Thereafter, the customer "confirms the payment (g)". If the payment is ok, "delivery service is activated (h)". If not, the customer must verify the payment data.

IS: Information System

Table 2: event log of purchase online process

Case ID	Activity	Resource	Date	Time	…
1	a	IS	23-05-2017	13:00:00	…
1	b	Jane	23-05-2017	13:10:00	…
2	a	IS	23-05-2017	14:00:00	…
1	c	Jane	25-05-2017	11:20:00	…
2	b	Ellen	25-05-2017	12:00:00	…
1	d	IS	25-05-2017	13:40:00	…
3	a	IS	25-05-2017	14:05:00	…
3	b	Adele	25-05-2017	15:00:00	…
1	g	Jane	26-05-2017	08:45:00	…
2	c	Ellen	26-05-2017	09:36:00	…
3	c	Adele	26-05-2017	11:00:00	…
2	e	IS	26-05-2017	11:15:00	…
3	f	Adele	26-05-2017	11:20:00	…
2	g	Ellen	26-05-2017	11:40:00	…
1	h	IS	27-05-2017	09:00:00	…
3	g	Adele	27-05-2017	09:30:00	…
3	h	IS	27-05-2017	10:00:00	…
2	h	IS	27-05-2017	11:20:00	

The data recorded in event logs present the execution history of one business process within an organization. A log case represents one process instance execution. The log represented in Table 2, records three different executions of the same process. Each process execution is called process instance, and referenced by an ID. The event log contains additional attributes, such as resource that executes the activity, date and time of the activity execution.

Definition 1: (Trace, Event log). Let A be a set of activities in some universe of activities. A trace σ ∈A∗ is a sequence of activities. An event log L∈T(A∗) is a multi-set of traces, i.e., an event log.

For instance, < a; b; c; d; g; h > is a trace that belongs to the event log in Table 2.

4.2. Log based relation

A log file is a set of traces. A trace can be defined as a sequence of events ordered chronologically and executed correctly. The execution order of activities in a process instance is of great importance. It helps to define dependency between activities and to capture all possible patterns encoded in the event log [1]. Based on the activities execution orders in traces, four ordering relations can derived from an event log: >L, →L, ||L and #L [22].

Definition 2: (Log-based ordering relations [22]) Let L ∈ M (A∗) be an event log over A. Let a; b∈A. The four ordering relations are precedence ($>_L$), causality ($\rightarrow_L$), parallelism ($||$) and choice ($\#_L$), defined as follows:

Precedence: $a >_L b$ if $\exists\sigma =< a1; a2; :::an >\in L$ and $1 \leq i \leq n - 1$ such that ai = a ^ ai+1 = b.

Causality: $a \rightarrow_L b$ if $a >_L b \wedge b >_L a$.

Parallelism: $a||_L b$ if $a >_L \wedge\ b >_L a$.

Choice: $a\#_L b$ if $a >_L b \wedge b >_L a$.

Based on the four log-based ordering relations, the discovered process model (figure 2) describes the behavior observed in the event log (table 2). The generation of the model is done by a discovery algorithm (e.g. [9,22]), and its representation by BPMN modeling language.

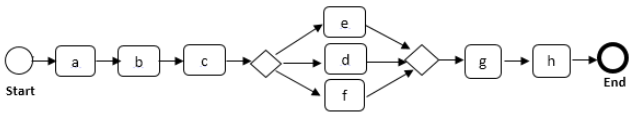


Figure 2: The process model corresponding to the event log in table 2

Event logs can be stored and exchanged using different forms of data source. MXML (Mining eXtensible Markup Language) is a standard notation for storing process attributes such as timestamps, resources and transaction types [23]. XES (eXtensible Event Stream) [24] is the MXML successor created to extend MXML.

4.3. Event log pre-processing

Logs are widely available in many applications, but the purpose of their creation and their level of details varies. To construct a configurable process model with meaningful behavioral patterns, the event logs must be pre-processed before using mining techniques.

- The elimination of confidential data is required before any data processing.
- The balance in the level of details in event logs is recommended. The generated process model has not to be highly detailed.
- The use of the same ontological concept for different sources of event logs.

5. Configurable process discovery approach

The construction of the process model can be done by merging models of process variants, which is complicated and error-prone especially when the number of variants is quite high. The approach we propose builds configurable process model from event log without merging exiting process models. It is based on event logs because of several reasons:

- Event logs are commonly available in Process Aware Information Systems (PAIS), such as : ERP, CRM and workflow management systems
- Business process models do not always exist. Therefore, techniques of merging cannot be applied.
- Event logs record process execution data exactly as it was executed in reality. The information recorded in the event logs is very useful for the business process design or configuration. For example, a priori process model does not present information like activity execution frequency, execution errors, and social behavior between users or services.

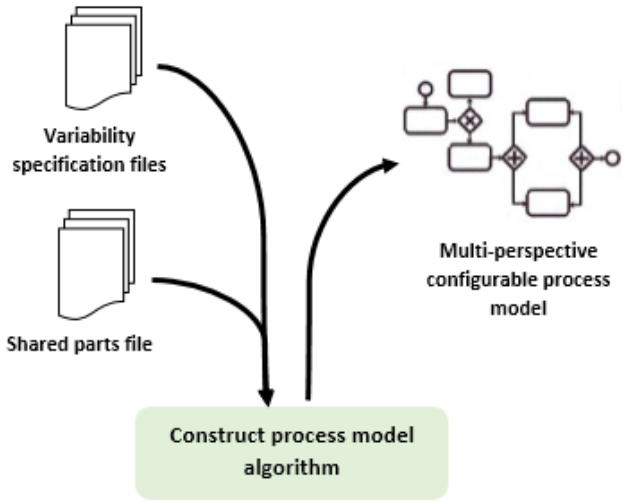


Figure 3: Construct configurable process model approach

In our approach (Figure 3), we discover configurable process model from event logs using two important files created directly from a collection of event logs: variability specification file and specification file of shared parts. The purpose behind the generation of these variability specification files is firstly to discover, in detailed manner, the variability of activity and resource and secondly to keep a record of variability for any forward configuration or update of the business process. The information recorded in the variability specification file represent variation points and variants of the configurable process while the specification file of shared parts represents the non-variable parts of the configurable process.

5.1. Framework architecture

The contribution presented in [11] proposes a variability discovery approach for business processes taking into account the

variability of resource perspective. In this paper, we shed light on our framework and its components. As well as, we describe the role of each component and we explain the interdependence between components. The discovery approach comes after, to create the configurable process model directly from event logs without using merging techniques.

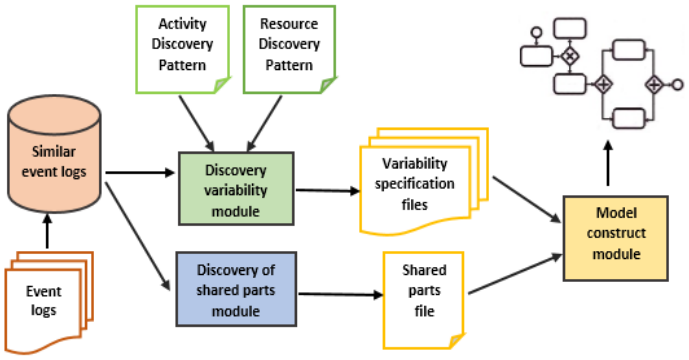


Figure 4: Architecture of the proposed framework [11]

The figure 4 presents the four components of the proposed framework:

- **Similar event logs:** It is a storage module of similar event logs collected and sorted by using existing techniques of clustering [9, 25, 26]. It takes as input a set of event logs and generates a set of pre-processed event logs.
- **Discovery variability module:** this module is important to discover the variability of different process perspectives namely activity, resource and data. It takes as input algorithms for activity and resource variability discovery and generates variability specification files for both activity and resource.
- **Discovery of shared parts module:** this module discovers the common parts shared by all process variants. It takes as input a set of similar event logs and generates specification file of shared parts of the process.
- **Model construction module:** this module constructs the configurable process model using the specification files of variability and shared parts.

The variability discovery module allows discovering the configurable fragments of the business process. It uses two algorithms to discover variability elements in detail. The first discovers the variability of activities [11] and the second discovers the variability of resources [23]. As output, the module generates i) a variability specification file, with variation points, variants, and variation point types for both activity and resource.

The shared parts discovery module discovers the common activities between all process variants. It generates ii) a specification file of common parts. This module allows discovering the non-configurable fragments of the business process.

Both variability discovery module and shared parts discovery module are essential for building the configurable model. The algorithm for configurable model creation uses i) and ii) for the configurable process model construction.

5.2. Discovery of variability

The discovery algorithm that we have introduced in [11] aims to provide an explicit variability discovery of activities. It discovers elements of variability such as variation points, type of variation point and variants. The algorithm uses discovery rules for each variability element [11]. These rules are used to define relations between activities and to construct the control flow process model.

Algorithm 1 : discovery of activity variability

```
1   Begin
2   Select the activity column
3   //Step 1: Discovery of variation point
4   For each activity do //From all activities of the process
5   For the direct activity successor do
6   Select successor1
7   // application of variation point discovery rule VP_rule
8   For each other_successor of the selected activity do
9     If successor1≠other_successor then
10              a variation point exists
11              save successor1
12              save other_successor
13    End if
14  //Step 2: discover of variants; application of var_rule
15   For each successor do
16    Define a variant
17  //Step3:name variation point as successor name
18  Name variation point=concatenation of all its
19  successors+vp as prefix
20  //Step4:define variation point type by application of
21  TVP_Parallel rule
22     If (successor1 is followed by other_successor)
23      and (other_successor is followed by successor1)
24     Then the point of variation is optional
25      End if
26     // application of TVP_Choice rule
27     If (successor1 is not followed by other_successor)
28     and (other_successor is not followed by successor1)
29     Then the point of variation is alternative
30     End if
31     End for
32  End for
33  End for
34  End
```

The discovery algorithm proceeds in different steps:

The algorithm discovers variation points based on the definition of activity successors from event logs. Then, it discovers the variants of variation points by defining the different successors identified in the previous step. After that, we concatenate the variants names to create the variation point name with add of the prefix “vp”. Finally, the algorithm discovers the type of variation point. For that, the log-based ordering relations [22] are used to discover relations between activities (choice or parallel).

The application of the discovery algorithm for activity variability generates variability specification file.

Definition 3: (variability specification file). Let A be a set of all activities of an event log L and a∈ A. Let VP, V be a two sets of finite activities, with VP ⊂ A, V ⊂ A and VP ⊔ V ⊆ A

Let o, a be the optional and alternative type of variation point as defined in [11], and let T= {t_o, t_a} be the type given to each variation point, respectively. Configurable fragment is defined as:

ConfigFrag(VP,T,V)= (a ∈VP, t∈T , (a_1,...a_n) ∈V n∈IN)

Variability specification file is a set of configurable fragments. It is defined as: ∑ ConfigFrag(VP,T,V)

This file contains details about variability elements of control-flow perspective. Different algorithms in our approach use this file for the discovery of resource variability and shared parts of the process variants.

5.3. Discovery of shared parts

To define shared parts of the process we implement algorithm based on the theory of sets in mathematics. It is defined as follow:

Let ET, EV and EC to be three sets of process activities. Where ET is the set of all process activities present in the event logs, EV is the set of activities present in the specification file of variability and EC the set common activities, we have: {E T = EV ⊔ EC } ⇔ {EV ⊂ ET , EC ⊂ ET , EV∩ EC =∅} ⇔{EC = ET\EV}

The algorithm uses the variability specification files and similar event logs. Which means that, the similarity between two activities name is maximum between their syntactic similarity and their linguistic similarity.

Algorithm 2 : Discovery of shared activities

```
1   init E(C) = ∅
2   begin
3       for each x ∈ E(T) do
4           if x not element of E(V) then
5           write x in E(C)
6           else continue
7           end if
8       end for
9     return E(C)
10  end
```

The algorithm parses the variability specification file of activities to discover the activities in common with all process variants, and then generates, as output, a specification file of common activities.

5.4. Creation of configurable process model

The discovery approach we propose constructs a configurable process model from event logs using the two files generated by the presented algorithm. The algorithm extracts the configurable fragment of the process model from the variability specification file of activities and extract the non-configurable fragment of the process model from the specification file of shared activities.

Algorithm 3 : Discovery of configurable model

```
1   Begin
2   Init Graph G
3   While E(C) ≠ ∅ do
4       for x ∈ E(C) do
5           if x ∈ E(V) then
6           G← create node x
7           define type of connector node
8           G ←create children nodes of  x successors
9           else
10          G ←create node
11          end if
12      end for
13    return graph G
14  end while
15  end
```

The algorithm parses the two generated specification files to construct the configurable model. It starts by selecting an element of shared parts and parses the variability specification file. If the element exists in the variability specification file, then the current element is a variation point, and the algorithm creates the corresponding node and its variants. If not, the activity is common of all process variants and the node is created.

6. Conclusion & future work

Many approaches were interested in control flow discovery, but few ones have been proposed for explicit variability discovery of different process perspectives. Given the importance of the variability for process reuse in business process management, the aim of our work is to propose a discovery approach of configurable process model from event logs, which provides a detailed information about variability elements in business process. In addition, the approach proposes a discovery algorithm for variability of other perspectives like activities and resources.

In the future work, we intend to work on the implementation of our approach with test results performed to evaluate its feasibility through different experiments. In addition, we will show the practical usefulness of our approach and publish a paper about the integration of resource variability in the discovered configurable process model.

References

[1] W.M.P. van der Aalst, Process mining: data science in action, 2nd edition Springer, 2016.

[2] M. Weske, Business Process process Management management concepts, languages, architectures. 2nd edition, Springer, 2005.

[3] M. Dumas, WM.P. Aalst, A.H.M. Hofstede, Process-Aware Information Systems Bridging People and software through process technology. John Wiley & sons, 2005.

[4] J. Becker, M. Kugeler, M. Rosemann, Process management: a guide for the design of business processes, Springer Berlin, Heidelberg, New York, 2003.

[5] Wil M. P. van der Aalst. "Workflow verification: Finding control-flow errors using petri-net-based techniques". In *Business Process Management, Models, Techniques, and Empirical Studies*, Springer-Verlag, 1806, 161-183, London, UK, 2000.

[6] R. Rychkova, S. Nurcan, "Towards adaptability and control for knowledge-intensive business processes: Declarative configurable process specifications", 1-10, IEEE Computer Society, 2011 https://doi.org/10.1109/HICSS.2011.452

[7] L. Seinturier, P. Merle, R Rouvoy, D. Romero, V. Schiavoni, J. B. Stefani. "A component-based middleware", 42(5), 559-583, May 2012.

[8] J.C.A.M. Buijs, B.F. van Dongen, W.M.P. van der Aalst., "Mining configurable process models from collections of event logs". In: Daniel F., Wang J., Weber B. (Ed.), Business Process Management. Lecture Notes in Computer Science, 8094, 33-48, Springer, Berlin, Heidelberg, 2013. https://doi.org/10.1007/978-3-642-40176-3_5

[9] Y.P.J.M. Van Oirschot, Using trace clustering for configurable process discovery explained by event log data, master thesis, University of technology Eindhoven, Netherlands, 2014.

[10] N. Assy, Automated support for configurable process models, PhD thesis, University Paris-SACLAY, France, 2015.

[11] R. Sikal, H. Sbai, L.Kjiri, "Configurable process mining: variability Discovery Approach", the 5th International Congress on Information Science and Technology,(IEEE CIST'18), Morocco, October 2018. https://doi.org/10.1109/CIST.2018.8596526

[12] R. Sikal, H. Sbai, L. Kjiri, "Promoting resource discovery in business process variability", proceedings of the 2nd international conference on Networking, Information System & Security, 48, 1-7. 2019. https://doi.org/10.1145/3320326.3320380
[13] F. Milani, M. Dumas, N. Ahmed, R. Matulevičius, "Decomposition driven consolidation of process models". In proceedings of the International Conference on Advanced Information Systems Engineering (CAiSE), 7908, 193-207. Springer, Berlin, Heidelberg, 2013.
[14] F. Gottshalk, W.M.P. van der Aalst, M.H. Jansen-Vullers, "Mining reference process models and their configurations". In R. Meersman, Z. Tari, P. Herrero (Ed.), On the Move to Meaningful Internet Systems: OTM 2008 Workshops, 5333, 263-272. Springer berlin Heidelberg, 2008.
[15] M. Rosamann, W.M.P. Van der Aalst, "A configurable reference modeling language". Information Systems, 32(1), 1-23, 2007. https://doi.org/10.1016/j.is.2005.05.003
[16] F. Gottshalk, W. M. P. van der Aalst, M. Jansen-Vullers, M. La Rosa, "Configurable Workflow Models", International Journal of Cooperative Information Systems, 17(2), 177-221. https://doi.org/10.1142/S0218843008001798
[17] H. Sbai, M. Fredj, L. Kjiri, "A pattern based methodology for evolution management in business process reuse". International Journal of Computer Science Issues, 11(1), 1-10, 2014.
[18] H. Sbai, L. El Faquih, M. Fredj, "A novel tool for configurable process evolution and service derivation". International Journal of Enterprise Information Systems (IJEIS), 15(2), 1-18, 2019.
[19] F. Gottshalk, W.M. P. van der Aalst, M. H. Jansen-Vullers, "Merging event-driven process chains". In R. Meersman, Z. Tari, P. Herrero (Ed.), On the Move to Meaningful Internet Systems: OTM 2008 Workshops, 5331, 418-426. Springer, berlin, Heidelberg, 2008. https://doi.org/10.1007/978-3-540-88871-0_28
[20] W. Derguech, S. Bhiri, E. Curry, "Designing business capability-aware configurable process models". Information Systems, 72, 77-94, 2017. https://doi.org/10.1016/j.is.2017.10.001
[21] M. Arriagada-Benítez, M. Sepúlveda, J. Munoz-Gama, J.C. A. M. Buijs, "Strategies to automatically derive a process model from a configurable process model based on event data". Applied sciences, 2017. https://doi.org/10.3390/app7101023
[22] W.M.P. van der Aalst, T. Weijters, L. Maruster. Workflow mining: Discovering process models from event logs. *16 (9)*. IEEE Transactions on Knowledge and Data Engineering, 2004. https://doi.org/10.1109/TKDE.2004.47
[23] B. F. Van Dongen. "A meta model for process mining data". In *Proceedings of the CAiSE WORKSHOPS*, 309-320, 2005.
[24] H.M.W. Verbeek, J.C.A.M. Buijs, B.F. van Dongen, W.M.P. van der Aalst. "XES, XESame, and ProM 6". In P. Soffer and E. Proper, editors, *Information Systems Evolution*, vol.72 of *Lecture Notes in Business Information Processing*, 60-75. Springer-Verlag, Berlin, 2010.
[25] M. De Leoni, W.M.P. van der Aalst, "A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs". In Information Systems, 56, 235-257. Oxford, UK, 2015. https://doi.org/10.1016/j.is.2015.07.003
[26] A. Appice, D. Malerba, A co-training strategy for multiple view clustering in process mining. IEEE Transaction on Services Computing, 9, 832-845, 2016. https://doi.org/10.1109/TSC.2015.2430327

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 191-200 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Biometric System Vulnerabilities: A Typology of Metadata

Abdou-Aziz Sobabe*, Tahirou Djara, Antoine Vianou

Laboratoire d'Electronique de Télécommunication et d'Informatique Appliquée (LETIA/EPAC) Université d'Abomey-Calavi, Benin, Institut d'Innovation Technologique (IITECH) Abomey-Calavi, Bénin

ARTICLE INFO

Article history:
Received: 19 September, 2019
Accepted: 13 January, 2020
Online: 22 January, 2020

Keywords:
Biometric vulnerabilities
Metadata
Area Under the Curve (AUC)

ABSTRACT

This study presents a root cause analysis of biometric vulnerabilities and provides a comprehensive typology of metadata in biometric adaptation. Although they are more reliable and secure than traditional authentication methods, biometric techniques are subject to vulnerabilities that pose challenges. Faced with the proliferation of cases of identity theft and fraud, biometrics is increasingly used to protect assets and people in several areas such as commercial, forensic and government applications. As a first step, a metadata analysis was performed. A focus has then been placed on their role in the fight against biometric vulnerabilities. Thus, the vulnerabilities studied have been classified into two main categories: intrinsic limits and adverse attacks. Finally, one of the scenarios considered was implemented, particularly the case of the combination of skin color with facial recognition. The implementation resulted in encouraging results with an Area Under the Curve (AUC) of 0.826 for the face system and 0.908 for the multimodal system.

1. Introduction

The fast-growing digital transformation is a process of fully integrating digital technologies into all of an organization's activities. In this digital world, the authentication stage is often considered as the weak link in computer security, because of the large number of identity theft cases. User authentication is an important challenge in securing assets. In reference to computer security, biometrics refers to the use of morphological, behavioral or biochemical characteristics to determine or verify the identity of a user.

Security is about "protecting" "assets" against all forms of "threats" [1]. Threat is the type of action that may be harmful in the absolute, while vulnerability (sometimes called a gap or breach) represents the level of exposure to the threat in a particular context. Finally, the countermeasure is the set of actions implemented to prevent the threat. The countermeasures to be implemented are not only technical solutions but also training and awareness measures for users, as well as a set of clearly defined rules. From the notions of threat, vulnerability and countermeasures, we can define the notion of risk as a danger that can be sustained at any time. The risk in terms of security is generally characterized by the following equation:

$$Risk = \frac{Threat * Vulnerabilities}{Countermeasures}$$

The security procedure can be represented by Figure 1 below.

As we can see, vulnerabilities are at the heart of the security procedure. Biometric authentication systems are positioned as the means of countermeasure put in place to ensure the protection of assets. But these systems face a number of challenges and vulnerabilities. Among the solutions to these challenges and vulnerability, the use of metadata figures prominently. In the following sections, we present an overview of biometrics metadata before addressing their role in addressing identified vulnerabilities.

2. Overview of some Non-Current Biometric Features

Biometric features are more reliable and secure for person recognition than traditional methods based on knowledge or possession. They can be morphological (palmprint, palm vein, (DNA, odor) or bioelectrical (electrocardiogram, electromyogram) [2–5].

2.1. Morphological Features

2.1.1. Hand and Finger Geometry

The human hand has some relatively stable features (e.g., length of fingers) which are peculiar (although not very distinctive) to an individual. They can therefore be used for biometric recognition purposes. The "measurement" of the hand is made up of several measures such as finger dimensions, joint

*Corresponding Author: Abdou-Aziz Sobabe, Email: azizsobabe@yahoo.fr

characteristics, palm and hand shape. Such an identification system studies on average 90 lines of the hand to recognize an individual. The first step of authentication is that of the scan. To do so, the person must put his hand on a platinum. The fingers must be properly placed. etc.), behavioral (voice, signature, gait, keystroke), biochemical. An infrared camera then takes an image from two different angles to obtain a three-dimensional reproduction of the hand. It normally takes no more than three seconds to read. The hand geometry systems have large physical size, so they cannot be easily embedded in existing security systems [6,7]. The hand geometry technology is used for applications such as access control, where one is cocerned about an individual trying to gain access by using someone else's card or PIN. In terms of advantages, the hand geometry authentication technique seems less intrusive than fingerprint registration. In addition, it requires the retention of smaller files compared to a database for comparing fingerprints. For example, scanning an image of a hand is only a space of about 15 bytes while the image of the fingerprint represents a file size of around 500 bytes or more. Finally, external factors such as the moisture of the skin and dirt on the hand do not prevent a good measurement. The same is true for burns and minor cuts. Compared to the fingerprint, this technique has a high rate of false positives. This means that family members with significant physical similarities can easily deceive the system. In addition, the shape of the hands is likely to change due to diseases related to old age, such as arthritis. Furthermore, the geometry of the hand requires a larger scanner than a fingerprint reader. As a result, its use becomes embarrassing when it comes to securing a small object such as a computer [8].

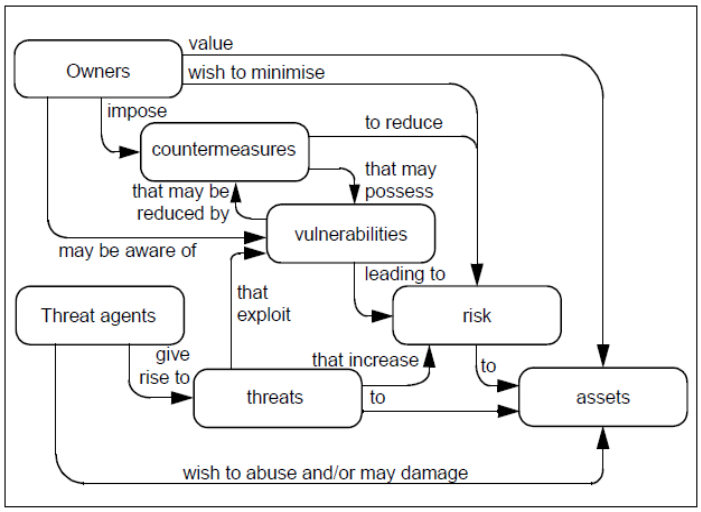


Figure 1: Security procedure [1]

2.1.2. Iris

Iris scan technology is one of the most reliable biometric recognition systems developed recently. The recognition of the iris is a biometric technique to recognize an individual by the observation of his iris. The iris is located in the aqueous humor, it is surrounded by the white of the eye, the pupil is located in its center, the cornea is in front of him and the crystalline behind. The iris corresponds to the colored part of the eye and it is this part that is used in biometrics. Indeed, the texture of the iris (i.e. the pattern of the iris), includes many features. Those most often used in biometrics are stripes, pits, and furrows. These elements of the iris remain fixed, they vary only very little during a lifetime. Each pattern determined by the chaotic processes during embryonic development is stable and unique (the probability of similarity is 1 in 10 power 72) [8,7]. In addition, the pattern of the iris is not genetic in contrast to the color of the eyes. So, two individuals, even if they are parents, can have the same iris color but never the same motive. That's why the iris makes it possible to distinguish the identical twins. Iris-based systems have a very low False Accept Rate (FAR) compared to other biometric traits. Nevertheless, it is important to note that iris systems have a high False Reject Rate (FRR) [6,9]. Iiris sample images quality have an impact on the accuracy of Visible Wavelength iris recognition systems. To deal with this problem, several iris image quality enhancement methods have been developed by researchers. Liu and Charrier [10] evaluated the performance of nine image enhancement approaches on Visible Wavelength iris biometric images. Experimental results showed that three of them (CLAHE, MEDIAN, and UNSHARP) can highly improve the system performance. In contrast, two other methods (SSR and WIENER) have a lower level of performance improvement.

2.1.3. Retina

The retina is the sensory membrane that lines the inner surface of the back of the eyeball. It's composed of several layers, including one that contains specialized cells called photoreceptors (https://www.allaboutvision.com/resources/retina.htm). The elements that distinguish two retinas are the veins that line them. The arrangement of these veins is stable and unique from one individual to another (from one eye to another). And the models that come from inherit the stability of this disposition. The difficulties involved in capturing the image of a retina are as much psychological as medical and technical. To obtain an image of a retina, it is necessary to illuminate the back of the eyeball by means of a light beam. This beam is of very low intensity so as not to disturb the user; it is safe and of lower intensity than on ophthalmic devices. A very precise camera system then comes to recover the image of the retina. Retina readers are available, and provide a very high level of security. After capturing an image of the retina, the reading device software cuts a ring around the fovea. In this ring, it locates the veins and their orientation. It then creates an "eye signature" that is used for retinal recognition. The operation itself is quite simple to describe but the algorithms remain relatively complex.

Retinal biometry provides a high level of recognition. This technology is well suited for high security applications (military sites, safes, etc.). The arrangement of the veins of the retina ensures a good reliability and a high barrier against fraud. [11].

2.1.4. Palmprint

Several authors have researched palm-based biometric recognition. These include [12,11,13]. Palm print recognition is a biometric authentication method based on the unique patterns of various characteristics in the palms of people’s hands. It is a relatively new biometric when compared to other biometric systems such as face, fingerprint and iris recognition systems. Palmprint is concerned with the inner surface of a hand. A palm is covered with the same kind of skin as the fingertips and it is larger

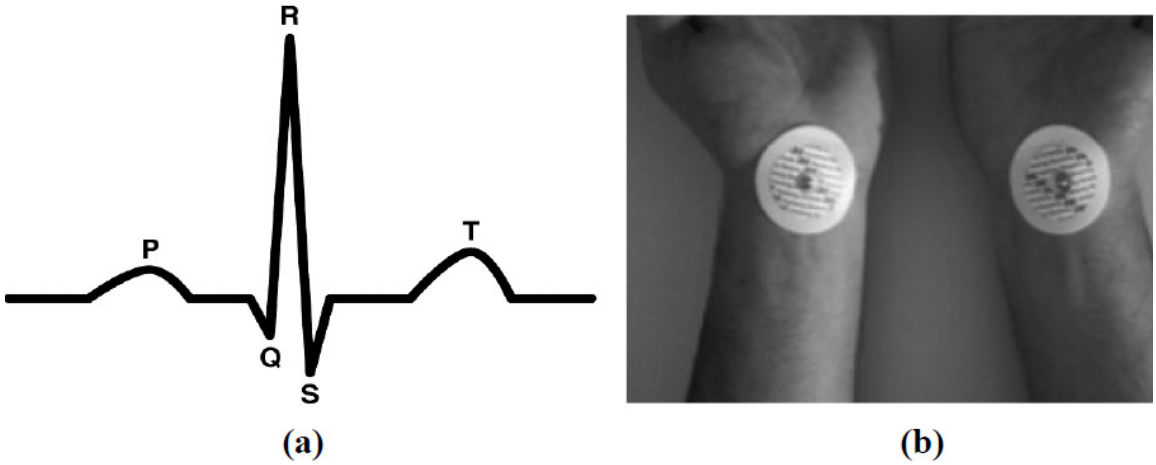


Figure 2: ECG biometrics: (a) ECG signal with regular rhythm (b) positioning of the electrodes on the forearms for ECG capture [14]

than a fingertip in size. Five features of a palmprint are usually used to uniquely identify a person. It is: (a) Geometry Features (such as width, length and area); (b) Principal Line Features (both location and form of principal lines); (c) Wrinkle Features (the thinner and more irregular lines); (d) Delta Point Features (the center of a delta-like region in the palmprint, usually located in the finger-root region) and (e) Minutiae Features (see fingerprint).

Other morphological traits are often used in biometric recognition. Those systems include Palm Vein, Ear and Facial Thermogram. More information can be found in [12,4].

2.2. Behavioral Features

2.2.1. Gait

Gait refers to the way a person walks, and is one of the few biometric traits that can be used to recognize people at a distance. Therefore, this trait is very appropriate in surveillance scenarios where the identity of an individual can be surreptitiously established [11]. Gait-based systems also offer the possibility of tracking an individual over an extended period of time. However, the gait of an individual is affected by several factors including the choice of footwear, nature of clothing, affliction of the legs, walking surface, etc. [9]. This type of biometric recognition system identifies individuals by their posture and the way they walk through unique characteristics that are perceivable at a low resolution, hard to conceal and non-invasive. [8].

2.2.2. Keystroke

Keystroke is a biometric solution "Software Only". It is applied to the password, which becomes much more difficult to "imitate". During the implementation of this technique, the user is asked to enter his password a dozen times in a row. Using an algorithm that exploits the support time on each key and the time between each key, the ten input is "averaged" to build a "typing profile" of the user that will serve as a reference. At the following accesses, the password entry will be coupled to a keying profile that will be compared to the reference profile. The access right is then granted according to the level of similarity of this profile with the reference. Depending on the degree of filtering that an administrator will have defined, this access will be more or less difficult. Whether additional security to an application, a Single Sign On, Intranet or Internet, in addition to security to a personal smart card, it is a solution, reliable, easy to implement and very competitive because without hardware. Its main advantages are related to rapid implementation for many users. It also significantly reduces the need for password changes and solicitation of IT services. As a major drawback, you must be in good condition before using the system at the risk of having your own password refused [8,11].

2.3. Bioelectrical Features or Hidden Features

2.3.1. Electrocardiogram

An electrocardiogram (ECG) is a test that studies how the heart works by measuring its electrical activity. At each heartbeat, an electrical pulse (or "wave") passes through the heart. This wave causes the heart muscle to contract so that it expels the blood from the heart [5]. The ECG is mainly used in clinical applications to diagnose cardiovascular disease. The ECG signal is characterized by the shape of its beats consisting of five typical waves, namely P, Q, R, S, and T or sometimes the U wave (Figure 2) [14]. ECG biometrics has been the subject of a number of studies [15–17]. The use of ECG in biometrics is relatively new. In fact, there are several biometric methods based on the ECG. There are approaches that are based on ECG analysis [18]. Others rely on the integration of analytic and appearance features extracted from ECG signals [19]. Several research studies confirm that the cardiac profile is specific to each individual and does not vary with age or even if the rhythm of the heart is racing after an effort or an emotion. Compared with other biometric modalities, the ECG is more universal and difficult to forge [20].

2.3.2. Electromyogram

Electromyogram (EMG) signals are bioelectrical signals recorded in the muscles. They provide various information on the state of the peripheral nerves. The EMG signal has several clinical applications. Its use as a hidden biometric modality can be particularly interesting. In this context, some recent experiments have been carried out [21,22,19]. In particular, this work has focused on the analysis of surface electromyography (SEMG) signals [23]. When acquiring these signals, individuals are encouraged to apply manual pressure of a constant intensity to a force probe for several seconds (Figure 3) [14]. The signal thus

obtained is analyzed in the spectral domain. Then, parameters are extracted such as the signal strength, the average frequency, the flattening coefficient and the asymmetry coefficient. Indeed, these parameters provide a device vector that we can use to characterize individuals.

In addition, it should be noted that magnetic resonance imaging (MRI) and X-rays are part of the hidden biometric techniques used for the authentication and identification of individuals [24].

3. Applications of Biometrics

The requirements related to the level of development of humanity and the security constraints require a rapid and reliable user authentication. Biometrics is an emerging field where technology improves man ability to identify a person [25]. The applications of biometrics can be divided into three main groups [26,13,27]:

- Commercial applications involve computer network access, electronic data security, e-commerce, Internet access, credit card, physical access control, cell phone, medical records management, remote study, etc One of the most popular technologies for ECG-based biometrics is the HeartID developed by researcher Foteini Agrafioti of the Canadian University of Toronto. This technology can be integrated into any electronic device, starting with smartphones, tablets or game consoles. The idea is to offer both an identification system to secure this type of device, but also to share them by activating user profiles opening on specific settings and contents. The sensors can be implanted so that the user only has to hold the terminal so that his heart rate is analyzed by the recognition algorithm, the identification taking about 1.2 seconds. A preliminary learning phase is necessary to record the electrocardiogram whose profile can then be stored in the terminal itself or on a database to which it will connect. Only requirement, the user must hold the terminal with both hands so that the signal can be captured because the electrocardiogram results from an electrical activity requiring two reference points on either side of the heart.

Another application of ECG-based biometrics is the Nymi bracelet developed by the Canadian company Bionym. This bracelet is able to detect the electrical signals emitted by the heart to control the identity of the wearer. Once identified, the owner can open his car, unlock his laptop or phone, or make payments such as with a credit card. With each electrical impulse, the heart will contract, and produce regular beats (between 60 and 80 per minute at rest). It is by analyzing this "cardiac signature" that the Nymi bracelet can identify its owner then, through a Bluetooth connection, interact with electronic devices to enter a PIN or unlock a user session.

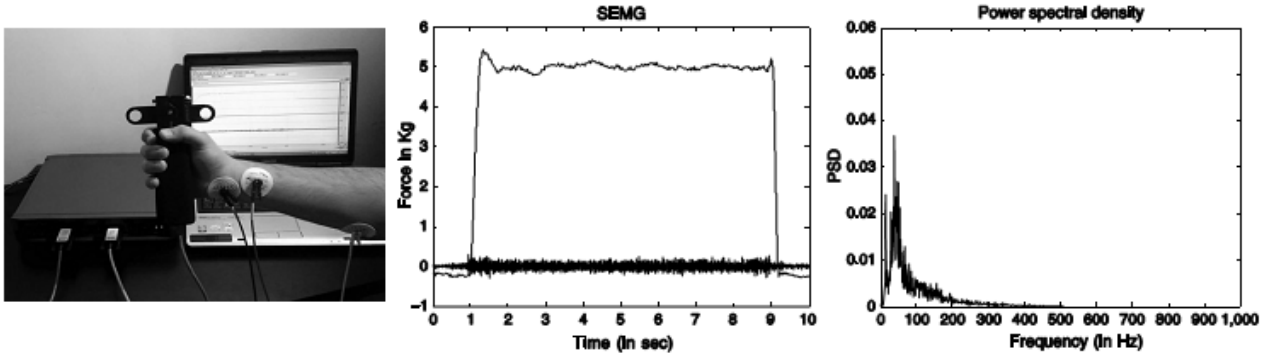
- Government applications include national identity card, driver's license, social security, border control, passport control, voter registration, etc. For example, the Government of Benin has created a Permanent Electronic Electoral List, the first draft of which was put online in January 2015. Since then, it has served as a basis for the various elections held in the country. The Schiphol Privium device at Amsterdam Airport uses an iris sensor to speed up the process of checking passports and visas.

- Forensic applications include body identification, criminal investigation, terrorist identification, parenthood determination, missing children etc. The Federal Bureau of Investigation (FBI) of the US Department of Justice utilizes the Integrated Automated Fingerprint Identification System (IAFIS), a 10-fingerprint matching system that captures rolled prints. In 2008, the Chinese Police adopted an Integrated Automated Biometric Identification System (IABIS) solution to allow forensic fingerprint examiners the ability to cross check inmate identities for possible matches within the database [28].

4. Typology of Metadata

Metadata or ancillary information is non-biometric data that is either combined with pure biometric data (fingerprint, face, iris, signature, etc.) or exploited by a system that makes biometric adaptation. The goal of adaptation is to provide ways to deal with the challenges of biometric systems.

In general, there are two main methods for adapting biometric systems, the first link to the user (soft biometrics) and the second link to the acquisition system. In the first case, there are a multitude of traits from the user that can be extracted and combined with pure biometric modalities. Jain et al. were among the first to explore this area. They did experiments on the following traits: Height, gender, ethnicity. Later, Dantcheva et al. [29] conducted a study on several soft biometric traits such as: age, skin color, hair color, eye color, marks, etc.



(a) (b) (c)

Figure 3: EMG biometrics: (a) Acquisition of an EMG signal (b) The intensity applied by the user and the relative EMG (c) EMG periodogram [14]

It should be noted that depending on the authors, certain traits are classified either as pure biometric modalities or as soft biometrics. This is the case of the gait which is sometimes considered as a behavioral modality [8] and sometimes as a trait of soft biometrics [29].

At the level of the acquisition system, adaptation can be done on the one hand at the sensor (volume, flash, etc.) and on the other hand taking into account the context of implementation (brightness, noise, etc.). Here, the metadata are not directly merged with the pure biometric data. The consideration is made at two levels. Either the system administrator makes adjustments to the various parameters such as volume, ambient brightness, flash, etc. On the other hand, the acquisition device is pre-sized to ensure automatic control of these different parameters. Such a device will be smart enough not to trigger for example a facial image when the brightness is not at a minimum required.

Several studies have focused on the role of metadata in improving biometric performance. In [30], the authors show the benefits of using gender, ethnicity and height information of the user in addition to fingerprint. The use of these soft biometric data leads to an improvement of approximately 5% over the main biometric system. [29] spoke of similar results which reduced the total error rate to 1.5% from 3.9% when body weight score is fused with fingerprint score. The table 1 below presents a typology of metadata in the biometric adaptation.

In the next section, special emphasis will be placed on the effect of metadata in the fight against biometric vulnerabilities.

5. Role of Metadata in Facing Biometric Challenges and Vulnerabilities

The challenges of biometric systems are essentially at the level of limits in terms of performance and acceptability. In addition, biometric authentication faces vulnerabilities that have been identified at eight levels on a generic biometric architecture [31]. Figure 4 shows the eight possible locations of attacks in a generic biometric system. [32] then listed a total of four different levels of biometric system vulnerabilities. These are intrinsic failure, administration issues, non-secure infrastructure, and biometric overtness. The last three levels have been grouped into adversary attacks from outside the system. These four levels of vulnerability were synthesized through a fish-bone model (see Figure 5).

5.1. Intrinsic Failure

An intrinsic failure is a biometric system security failure that generates an incorrect decision taken in a native way ie without without the intervention of an external element. It can occur even in the absence of an explicit effort by an attacker to bypass the system. This type of failure is also known as "zero-effort-attack". A biometric verification system can make two types of errors in decision making. These are false acceptances and false rejections. In the first case, false acceptances are usually caused by a lack of uniqueness in the biometric trait, which can lead to a great similarity between the characteristics of different users. At this level, adding metadata can be effective in combating this problem. Among the most probable cases, there are those of two identical twins or brothers who have a perfect similarity of the face for example. Taking into account soft biometric data such as marks and make-up will make it possible to distinguish between two individuals even if they have exactly the same face because the face recognition algorithms do not take into account these accessory data.

A legitimate user may be falsely rejected by the biometric system because of the large differences between the user's stored model and the input biometric feature sets. These intra-user variations may be due to an incorrect interaction of the user with the biometric system. This is the case, for example, of the effect of twists on a fingerprint image or the effect of the changes of pauses on a face image. Metadata do not provide a specific solution for these cases of intra-user variations. But increasingly, the biometric verification algorithms take into account these issues. [33] have proposed a contactless fingerprint recognition algorithm that addresses this issue. On another level, it should be noted that the false rejections can be due to noise introduced at the sensor such as, for example, the residual traces left on a fingerprint sensor. Metadata does not provide a solution to this limit. Multimodal

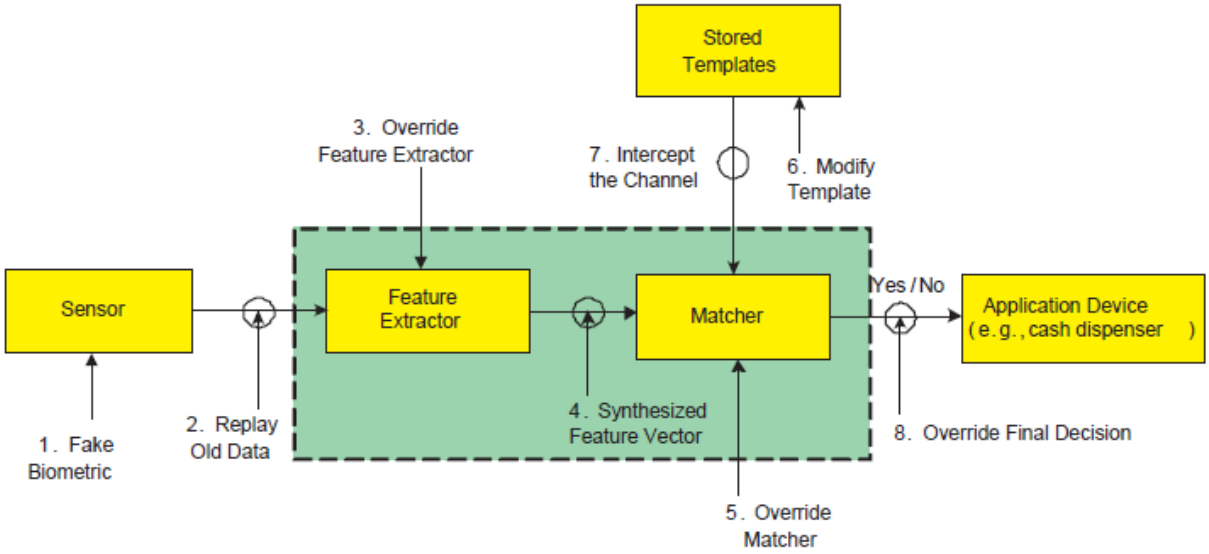


Figure 4: Possible locations of attacks in a generic biometric system [32]

biometrics offers an alternative to false rejects. Indeed, when one of the modalities is affected of noise, the system is able to ignore it and pass to the following modality.

For the limits related to the detection technology, the exploited metadata can be the volume (case of the voice modality) or the flash (case of the face, the contactless fingerprint, etc.). Finally talking about limits in the context of implementation, parameters such as brightness, noise, and temperature are used as metadata to deal with intrinsic errors.

5.2. Adversary Attacks

An adversary intentionally launch an attack on the biometric system by exploiting a multitude of loopholes in the system design. The three possible forms of attack are: Administration attacks, non-secure infrastructure attacks, and biometric overtness attacks.

5.2.1. Administration Attack

This attack is still called insider attack. It usually comes from malicious, familiar people who exploit vulnerabilities created by poor biometric system administration. These include (i) the integrity of the registration process (for example, the validity of credentials submitted during registration), (ii) collusion (or coercion) between the opponent and the system administrator or a legitimate user and (iii) abuse of exception handling procedures. Under these conditions, the use of metadata will not provide a response to this type of attack in the absolute. On the other hand, the additional use of soft biometric traits can make the system more complex and require more effort on the part of the attacker.

5.2.2. Non-Secure Infrastructure

The infrastructure of a biometric system includes the hardware, software and communication channels between the different modules. An opponent can manipulate the biometric infrastructure in several ways, which can lead to eight security breaches as described below.

First level:

Falsified biometric data: A reproduction of the biometric data used will be presented to the biometric sensor. In the case of authentication by the fingerprint, the attacker can present a false finger facing a contact-based sensor or just have the image of a finger facing a contactless sensor. It should already be noted that this form of attack is more common with the fingerprint which is a widely used modality with a high discriminatory power. On the other hand with other modalities such as face, retina, and signature, it is more difficult to present a falsified data. Moreover, if the biometric system integrates a soft biometric feature to be combined with pure biometric modality, this form of attack becomes more complicated to implement. Take for example a case of fingerprint authentication associated with the height of the user. Even if the fingerprint is falsified, the impostor will be stucked during the enrollment of the height. Considering another case of face authentication combined with the skin color automatically extracted from the facial image, an impostor will have a hard time tampering with a face while taking into account the specificities and requirements related to the detection of the skin color.

Second level:

Transmission of intercepted biometric data: Here, the attacker plays back an old biometric data stored in the system without passing through the biometric sensor. This is the case with the presentation of an old copy of the fingerprint image. Since the attacker bypasses the biometric sensor by providing the system with an old recorded data, the metadata will have no effect against this form of attack.

Third level:

Attack on the feature extraction module: This module could be replaced by a Trojan horse so as to produce information chosen by the attacker. The legitimate user does not realize that this module has been corrupted and provided information in accordance with the hacker's instructions. The feature extraction module being compromised by the hacker, the metadata will not be effective against this kind of attack.

Fourth level:

Alteration of extracted features: After data is obtained by the feature Extraction Module, it is altered or even replaced by other data defined by the attacker. For non-secure infrastructure attacks, from the second level of attack to the eighth level, we are in situations where the biometric system is corrupted and will only provide responses according to the intent of the hacker. Metadata will not be effective in these contexts.

Fifth level:

The matching module is replaced by a malicious module: This module could be replaced by a Trojan horse to artificially produce high or low scores.

Sixth level:

Database corruption: The database of biometric templates is available locally, remotely or distributed on multiple servers. In this type of attack, the attacker modifies one or more models to allow an impostor or even to prevent a legitimate user from accessing it.

Table 1: Typology of metadata in biometric adaptation

Level of adaptation	User	Acquisition system	
Type of adaptation	Soft biometric	Sensor	Context
Example	Marks; Make-up; Age; Weight	Volume (microphone); Flash (camera)	Brightness, noise, temperature
Exploited data	Metadata		
Training need	Without training (immutable)		After training

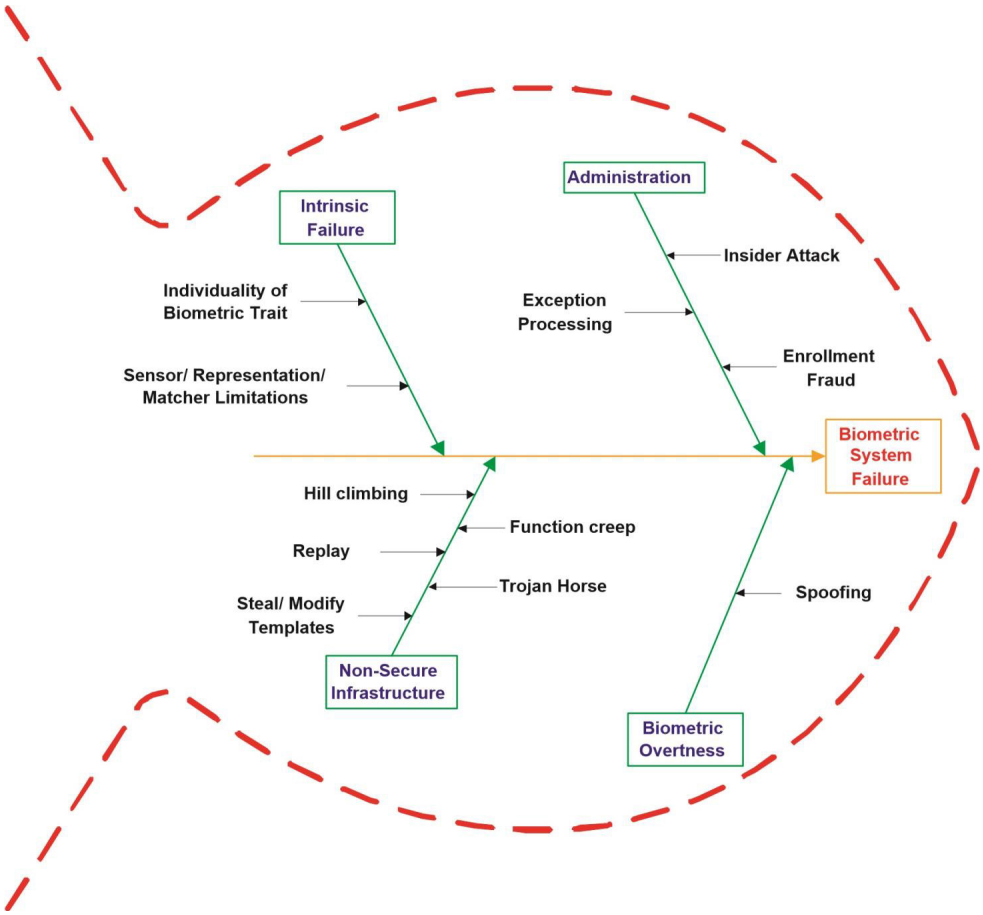


Figure 5: Fish-bone model for categorizing biometric system vulnerabilities [32]

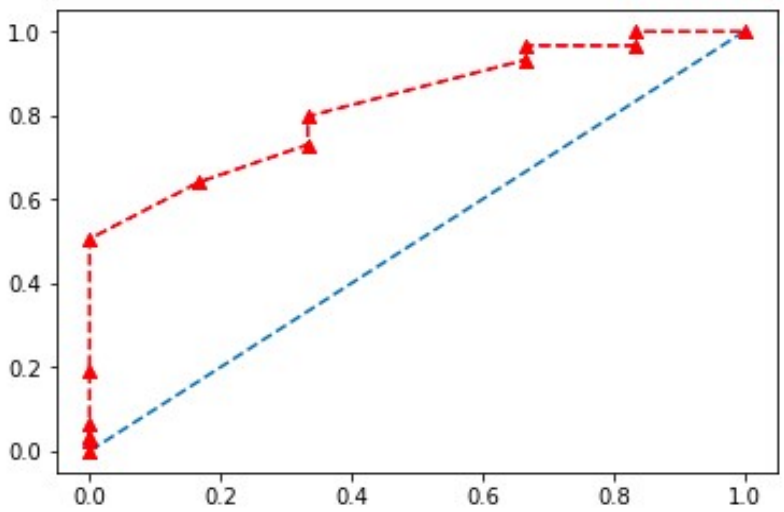


Figure 6: ROC curve with AUC: 0.826

Seventh level:

Attack on the channel between the database and the matching module: In this type of attack, the models are altered on the transmission link connecting the model base and the matching module.

Eighth level:

Alteration of decisions (accepted or rejected): This type of attack alters the boolean decision (yes or no) made by the decision module. The danger of this attack is high because even if the system is robust in terms of performance, it was rendered useless by this type of attack.

5.2.3. Biometric Overtness

An attacker may discreetly acquire the biometric characteristics of a legitimate user and use them to create an artificial production of the biometric feature (fingerprints taken from the surface of a contact-based sensor). Therefore, if the biometric system is not able to distinguish between a live biometric presentation and an artificial parody, an opponent can circumvent the system by presenting falsified traits. This case perfectly matches the first level of attack on non-secure infrastructure. The analysis performed at this level is also valid here. It should be noted that soft biometrics represents a response to this form of attack for two reasons. Indeed, the user is obliged to appear in person in front of the data acquisition device or this data is automatically extracted from a pure biometric modality having a number of characteristics hardly respectable by the artificial productions.

The table 2 below summarizes for each vulnerability category the existence of effects as well as the possible actions of the metadata. Examples of usable metadata have been listed at different levels in order to provide more probable scenarios of integration of multi-biometric metadata using a variety of disparate fusions of traits.

6. Skin Color as a Solution to Overtness in Face Authentication

6.1. Materials and Methods

The objective of this section is to present a practical case of adding metadata to a pure biometric modality in order to fight against certain vulnerabilities. The materials and methods used are among the most recent technologies in progress. The implementation of the proposed system was carried out under Python framework with the OpenCV package (https://sourceforge.net/projects/opencvlibrary/). Two stages are to be considered in accordance with the two modalities that are the face and the skin color (of the face). First, we performed face recognition using the LBPH (Local Binary Pattern Histogram) algorithm [34]. Then, for authentication by skin color, the Haar Cascade algorithm [34] made it possible to detect the face, while the extraction of the dominant skin colors was carried out automatically using the algorithm of the K-means [35]. Only one database was used; it is the Caltech faces database (http://www.vision.caltech.edu/html-files/archive.html). This database contains 450 images of faces of size 896x592 in JPEG format for 27 unique people with a variable number of images. For experimental purposes, we have resized the database to 19 people with 17 images each. 12 images were used to form the model and 5 images were used for the tests.

6.2. Experimental Results

After implementing these two modalities, the matching scores obtained were merged using the weighted sum method. A weight of 0.9 was assigned to the modality of pure biometrics (face) while the modality of soft biometrics (skin color) received a weight of 0.1. The experimental results are satisfactory with a performance of 82.6% for the facial system and 90.8% for the multimodal system. Figures 6 and 7 respectively show the results (ROC curve) of the face alone and of the face combined with the skin color.

6.3. Discussion

The inclusion of metadata in face authentication allows on the one hand to improve recognition performance and on the other hand to make the system less vulnerable. The calculation of the processing time for the two modalities gave the following results. For the face, the average processing time (for 15 images) is 1.24 s. The average time for skin color is 7.21 s. The fusion of the facial and skin color scores takes an average time of 1.52 s (always for the 15 images). It should be noted that the total processing time of the multimodal system increases due to two factors, namely the extraction time of the skin color and the time taken to merge the two scores. This confirms once again that each biometric authentication system has its advantages and limits. These are the application cases and the implementation conditions which generally determine the choices to be made.

7. Conclusion and Future Work

This paper presented a detailed typology of metadata on multi-biometric system vulnerabilities. We have provided an overview of some non-current biometric sensing systems before dealing with the applications and privacy issues. Furthermore, we discussed the categorization of biometric metadata, namely user level (ie soft

Table 2: Typology of metadata on biometric system vulnerabilities

Vulnerability category	Intrinsic failure	Adversary attacks		
		Administration Attack	Non-secure infrastructure	Biometric overtness
Existence of metadata effect	Yes (except the case of false rejects)	Partial	Yes (1 case out of 8)	Yes
Possible actions	1- Reduce false acceptances 2- Limit FTE and FTA	Make the system harder to attack	Prevent attack by presenting falsified biometric data	Prevent attack by presenting falsified biometric data
Example of metadata	1- Marks (tattoos, scars, etc.); make-up 2- Volume, flash; Brightness, noise, temperature	Weight, body shapes, etc.	Height, eye color, etc.	Gait, gender, skin color, etc.

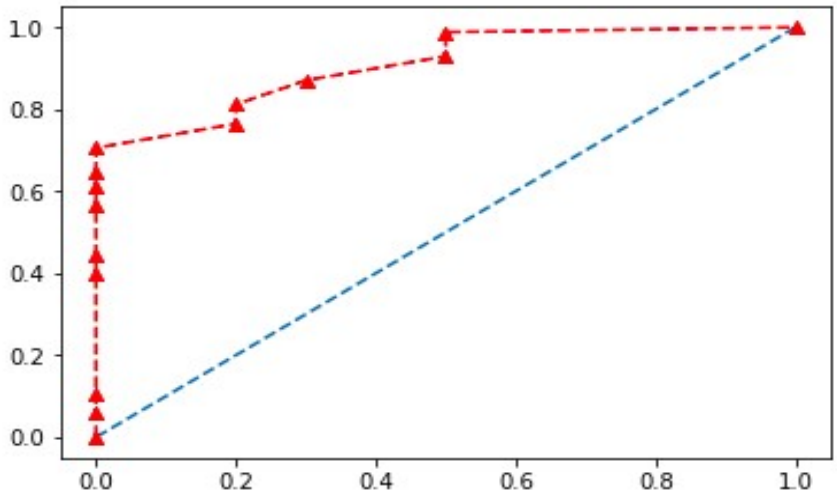


Figure 7: ROC curve with AUC: 0.908

biometric) and acquisition system level (ie sensor and acquisition context). To finish, a practical case of adding metadata to a pure biometric modality have been presented in order to fight against face recognition vulnerabilities.

Contactless fingerprint authentication [33] is a promising technique but one that is potentially vulnerable on several levels. For future work, we will explore the role of metadata in addressing contactless fingerprint vulnerabilities.

References

[1] [Common Criteria, 2012] Common Criteria for Information Technology Security Evaluation ; Part 1: Introduction and general model ; September 2012 ; Version 3.1 ; Revision 4

[2] M. Nageshkumar, P.K. Mahesh, and M.N. Shanmukha Swamy, "An efficient secure multimodal biometric fusion using palmprint and face image," International Journal of Computer Science Issues, Vol. 2, pp 49-53, 2009, arXiv:0909.2373v1.

[3] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," Elsevier, Pattern Recognition 38, pp 2270-2285, 2005, DOI:10.1016/j.patcog.2005.01.012.

[4] R. Gad, A. El-Sayed, N. El-Fishawy, and M. Zorkany, "Multi-biometric systems: A state of the art survey and research directions," International Journal of Advanced Computer Science and Applications, Vol. 6, No. 6, pp. 128-138, 2015, DOI : 10.14569/IJACSA.2015.060618.

[5] S. N. Abbas, M. Abo–Zahhad, S. M. Ahmed, M. Farrag, "Heart-ID: human identity recognition using heart sounds based on modifying melfrequency cepstral features," © The Institution of Engineering and Technology, IET Biom., Vol. 5, Iss. 4, pp. 284-296, 2016, doi: 10.1049/iet-bmt.2015.0033.

[6] A.K. Jain and A. Kumar, "Biometric recognition: An overview," Chapter 3, © Springer Science + Business Media B.V., pp 49-79, 2012, DOI 10.1007/978-94-007-3892-8_3.

[7] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar, Handbook of Fingerprint Recognition, © Springer-Verlag London Limited, 2009.

[8] M. O. Oloyede and G.P. Hancke, "Unimodal and multimodal biometric sensing systems: A review," IEEE Access, Volume 4, pp 7532-7555, 2016, Digital Object Identifier: 10.1109/ACCESS.2016.2614720.

[9] A. Ross, K. Nandakumar, and A.K. Jain, Handbook of Multibiometrics, Springer, New York, USA, 1st edition, 2006.

[10] X. Liu and C. Charrier, "Can image quality enhancement methods improve the performance of biometric systems for degraded visible wavelength iris images?," BioSMART 2019 Proceedings, 3rd International Conference on Bio-engineering for Smart Technologies, Paris, 24th-26th April pp. 119-122, 2019, 978-1-7281-3578-6/19/$31.00 ©2019 IEEE.

[11] H. AlMahafzah and M.Z. AlRwashdeh, "A survey of multibiometric systems," International Journal of Computer Application, volume 43 No 15 April, pp 36-43, 2012, DOI : 10.5120/6182-8612.

[12] D. Cherifi, R. Kaddari, H. Zair, and A. Nait-Ali, "Infrared face recognition using neural networks and HOG-SVM," BioSMART 2019 Proceedings, 3rd International Conference on Bio-engineering for Smart Technologies, Paris, 24th-26th April, pp. 132-136, 2019, ©2019 IEEE, 978-1-7281-3578-6/19/$31.00.

[13] A.K. Jain, P.J. Flynn, and A. Ross, Handbook of Biometrics, © Springer Science + Business Media, LLC, 2008.

[14] H. Toufik, "Reconnaissance Biométrique Multimodale basée sur la fusion en score de deux modalités biométriques : l'empreinte digitale et la signature manuscrite cursive en ligne," Ph.D Thesis, Badji Mokhtar University-Annaba, 2016.

[15] Q. Zhang, "Wavelets networks: The radial structure and an efficient initialization procedure," Technical Report of Linköping University, 1992, LITH-ISY-I-1423.

[16] J. Sjöberg , Q. Zhang, L. Ljung, A. Benveniste, B. Deylon, P.Y. Glorennec, H. Hjalmarson, and A. Juditsky, "Nonlinear black-box modeling in system identification: Unified overview," Automatica, Vol.31, no.12, pp.1691-1724, 1995.

[17] A. Juditsky, "Wavelet estimators: Adapting to unknown smoothness," Technical Report IRISA, June 1994.

[18] Q. Zhang, "Using Wavelet Network in Nonparametric Estimation," Publication Interne 833, IRISA, June 1994.

[19] Q. H. Zhang, "Using Wavelet Network in Nonparametric Estimation," IEEE Trans.Neural Networks, Vol.8, pp.227-236, 1997.

[20] S. Chantaf, "Biométrie par signaux physiologiques," Ph.D Thesis, Université Paris-Est, 2011.

[21] J. Zhang, G. G. Walter, Y. Miao, and W. N. W .Lee, "Wavelet Neural Networks for function learning" IEEE Trans. signal process, Vol.43, No.6, pp.1485-1497, June 1995.

[22] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet transform," IEEE Trans. Pattern Anal. Machine Intell., Vol.11, pp.674-693, July 1989.

[23] N. Draper and H. Smith, "Applied regression analysis, Series in Probability and Mathematical Statistics," Wiley, 1981, Second edition.

[24] A. Nait-ali, "Hidden biometrics: towards using biosignals and biomedical images for security applications," 7th international Workshop on Systems, Signal Processing and their Applications, Invited paper, Tipaza, pp. 352–356, 2011, DOI: 10.1109/WOSSPA.2011.5931509.

[25] L. Allano, La Biométrie multimodale : stratégies de fusion de scores et mesures de dépendance appliquées aux bases de personnes virtuelles, PhD thesis, Institut National des Télécommunications dans le cadre de l'école doctorale SITEVRY en co-accréditation avec l'Université d'Evry-Val d'Essonne, 2009.

[26] A.K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Transactions on circuits and systems for video technology, vol. 14, no. 1, January, pp 4-20, 2004, DOI: 10.1109/TCSVT.2003.818349.

[27] S. Guerfi, "Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D," PhD thesis, Université d'Evry-Val d'Essonne, 2008.

[28] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(1), January 1994, pp 66–75, DOI: 10.1109/34.273716.

[29] A. Dantcheva, C. Velardo, A. D'Angelo, J. L. Dugelay, Bag of Soft Biometrics for Person Identication : New trends and challenges, Multimed Tools Appl 51, © Springer Science+Business Media, LLC 2010, pp. 739–777, 2011, https://doi.org/10.1007/s11042-010-0635-7.

[30] A.K. Jain, K. Nandakumar, X. Lu, U. Park, "Integrating faces, fingerprints, and soft biometric traits for user recognition," Proceedings of Biometric Authentication Workshop, LNCS 3087, Prague, pp. 259-269, May 2004, DOI:10.1007/978-3-540-25976-3_24.

[31] M. El-Abed, C. Charrier, "Evaluation of biometric systems," chapter 7. In: New Trends and Developments in Biometrics, pp. 149–169, 2012, https://doi.org/10.5772/52084.

[32] A.K. Jain, K. Nandakumar, A. Nagar, "Biometric template security," in EURASIP Journal on Advances in Signal Processing, Special Issue on Biometrics, pp 1-20, January 2008, DOI:10.1155/2008/579416.

[33] T. Djara, M.K. Assogba, and A. Vianou, "A contactless fingerprint verification method using a minutiae matching technique," International Journal of Computer Vision and Image Processing, Volume 6, Issue 1, pp 12-27, January-June 2016, DOI: 10.4018/IJCVIP.2016010102.

[34] L. Dinalankara, "Face detection and face recognition using open computer vision classifiers", Robotic Visual Perception and Autonomy, Faculty of Science and Engineering, Plymouth University, august 4, 2017.

[35] C. C. Aggarwal, C. K. Reddy, Data clustering : algorithms and applications, Taylor & Francis Group LLC, 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 201-211 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

An Algorithm for Automatic Measurement of KI-67 Proliferation Index in Digital Images of Breast Tissue

Guillermo Kemper[*,1], David Atencia[1], Ivan Ortega[1], Roberto Kemper[2], Alejandro Yabar[3]

[1]*Faculty of Engineering and Architecture, School of Electronic Engineering, University of San Martín de Porres, Lima 12, Peru.*

[2]*Institute of Pathology and Molecular Biology Arias Stella, Lima 11, Peru.*

[3]*Department of Anatomy Pathology, National Hospital Edgardo Rebagliati Martins, Lima 11, Peru.*

ARTICLE INFO

Article history:
Received: 17 October, 2019
Accepted: 24 December, 2019
Online: 22 January, 2020

Keywords:
Breast Cancer
Image Processing,
Immunohistochemistry
KI-67

ABSTRACT

This paper proposes an algorithm aimed at quantifying the expression of KI-67 protein in digital images of breast biopsy tissue samples obtained through an optical microscope. The algorithm allows to obtain a report on the quantity of non-proliferating and proliferating cells through the detection and quantification of KI-67. The sample analysis via software aims to reduce the level of subjectivity in the diagnosis of diseases such as breast cancer. The algorithm proposed involves the application of statistical image processing techniques, adaptive thresholds, object segmentation and color filtering. A method of analysis and quantification of overlapping cells is also proposed to improve the efficiency of the algorithm. The results of the method proposed were satisfactory, as they were highly correlated with those obtained by visual inspection by pathologists.

1. Introduction

Cancer is a public health problem worldwide, as evidenced by its high rates of incidence and mortality. In Peru, for example, the Lima Cancer Registry reported, in 2011 alone, about 34,000 new cases of cancer [1]. Likewise, according to the Peruvian Epidemiological Surveillance System, in the period 2006-2011, breast cancer was the second most common among Peruvian women [2], in order to choose the correct therapeutic treatment for breast cancer, guidelines often combine conventional predictors to estimate relapse/mortality risk. [3][4]

In medical research, some biomarkers have been validated for the detection and estimation of the disease progression, including breast cancer. Among them are the estrogen receptor (ER), the progesterone receptor (PR), HER2 and KI-67 [3][4]. Through a quantitative analysis, KI-67 provides a reliable estimate of the proliferative activity in a tissue. Its application to biopsies in patients with breast cancer gives doctors some evidence to treat the disease [4][5].

KI-67 is a protein that is present in all cell cycle phases (G1, S, G2, M), except for the resting phase (G0); its presence is, therefore, quite intense and large in neoplastic proliferations [3]. For this reason, the quantification of KI-67-positive cells provides valuable information to determine the magnitude of the proliferative activity.

To prove the existence of protein KI-67 in a breast biopsy tissue sample, a specific anti-protein is used with a chromogen additive allowing visualization, denominated MIB-1, this cell staining is made in a process called Immunohistochemistry (IHC). Finally, the specialist analyses the sample (in a subjective way) by eyeballing inspection through a microscope to give a report to the patient on the proliferative activity of KI-67-positive cells. According to the St. Gallen Consensus (S. Bustreo, S. Osella-Abate, P. Cassoni et al, 2016), the KI-67 index is one of the most important prognostic markers used by oncologists to determine the treatment of ER-positive breast cancer patients [6]. The number of proliferating cells in proportion to the total number of cells in a particular tissue area, called hereinafter 'proliferative index' (PI), is divided into 2 levels of prediction for Ki-67: Luminal A, if the proliferative index is less than 20%; and Luminal B, if it is over or equivalent to 20% [6].

However, subjective evaluation can lead to errors in diagnosis due to several factors: room lightning, stress, fatigue, concentration, visual ability, experience, etc. They often cause that the results provided by pathologists are uncorrelated and have high

[*]Corresponding Author: Guillermo Kemper, Email: gkemperv@usmp.pe

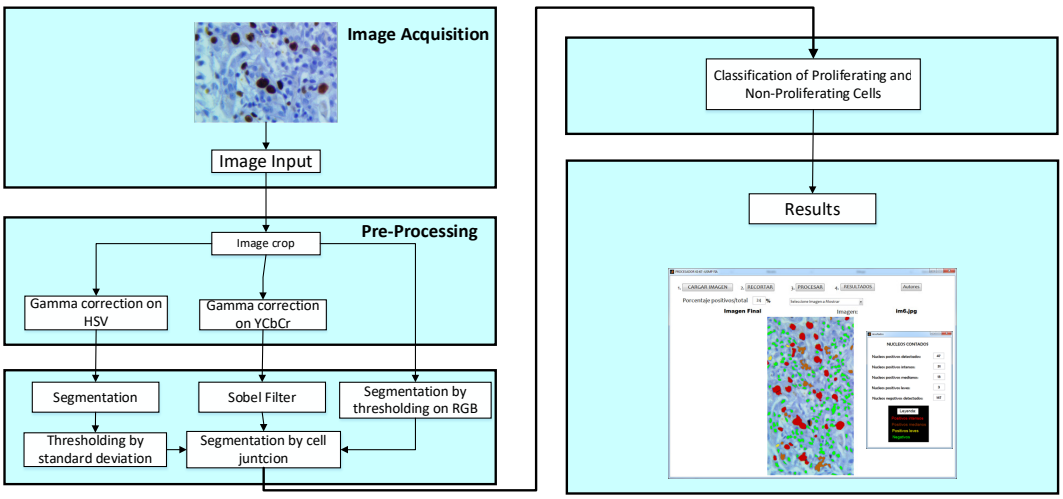


Fig. 1. Block diagram of the algorithm proposed.

variance [7][8], another factor that influences is the heterogenity of the stained sample [9]. Possible solutions to this issue have been proposed in previous research, but the main problems affecting cell detection in similar software remain: imperfection in collecting data such as noise, colour distortion and deformity in histopathological material at the stage of sample pigmentation, therefore a good calibration of the staining machine between IHC test should be made to avoid difference in pigmentation, this create a inter-laboratory variability on identical set of tumors [10].

In [11], for example, the results of the quantification algorithm were successful in analysing KI-67 with different image processing techniques such as thresholding, segmentation and labelling. However, the use of the ultra-erosion technique damaged valuable information for the evaluation of the sample.

In [12], they implement an algorithm based on mathematical description of cell morphology added with Support Vector Machine (SVM), obtaining great accordance between the digital image analysis and the "eyeballing" method.

On the other hand, M. Bouzid et al. show in [13] an interesting proposal with the use of special hardware: a multispectral camera with programmable light source. However, it increased the cost and availability of the solution. F. Kabir and N. Yusoff used in [14] the method of Random Forest (RF), consisting of predictive tree-like structures, each independent of the other, that obtain their data from different random vectors, but same distribution. RF has proved to perform better than other machine learning techniques, such as neural networks, support vector machines and k-nearest neighbors. The results of the tests, however, showed an accuracy of 72% in differentiating between malignant and benign tissue.

In this paper, an alternative solution is proposed on the basis of a digital image processing algorithm designed to quantify and provide more objective results regarding the level of cell proliferation based on marker KI-67. The algorithm involves statistical image processing techniques [15], adaptive thresholds [15][16], object segmentation and colour filtering [17]. Similarly, it proposes an overlapping cell analysis and quantification method to improve the efficiency of cell detection. These techniques have been used together to obtain satisfactory results based on the problem. The results were indeed satisfactory, as presented at the end of this document.

In the following sections, the processing stages of the algorithm proposed in this paper will be described.

2. Description of the proposed algorithm

The proposed algorithm involves different processing parts and stages, as shown in Fig. 1. The scheme proposed contains the stages of image acquisition, pre-processing, segmentation, classification and counting. These steps will be described below.

2.1. Image acquisition

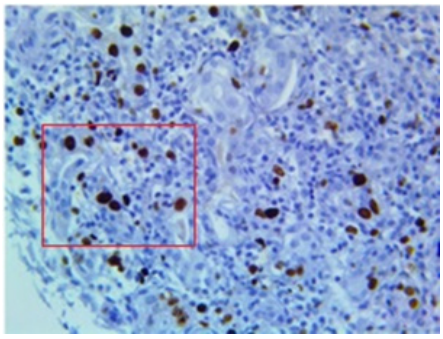
Before being digitalised, the biopsy sample goes through an IHC process, which mixes the laminated sample with chemicals and special antibodies to achieve proper homogenous staining.
Currently, the procedure is performed automatically using the Ventana BenchMark XT stainer in line with the producer's manual. After staining, the sample is placed under an Olympus BX53 digital microscope with a 40x magnifying lens, 50 images were chosen to study, all were given by chief of the department of anatomy pathology of Hospital Edgardo Rebagliati Martins, all samples were chosen randomly. The images of the biopsy sample (pre-treated by IHC) are obtained from a digital camera and an optical microscope. In this case, camera resolution was *M0*=2748 rows by *N0*=3584 columns with 40x zoom. The colour format of the original image (Fig. 2a) is RGB (Red, Green and Blue) called $I(x, y, z)$ with 12 bits per primary colour component (36 bits/pixel). The primary components can be expressed as $F_R(x, y), F_G(x, y)$ and $F_B(x, y)$.

All the parameters used in this paper were chosen taking in account the procedure mentioned previously to convert the biopsy sample into a digital image.

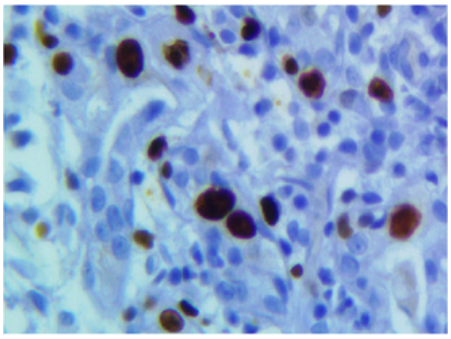
2.1.1. Choosing the area of interest

The image can be cropped to select an area of interest. This procedure grants the pathologist the freedom to choose the area to analyse according to their own criteria and preference.
The selected area will have a resolution of *M* rows by *N* columns (this values may vary depending on the area selected by the

specialist). The format is also RGB with primary components $\boldsymbol{I_R(x,y)}, \boldsymbol{I_G(x,y)}$ and $\boldsymbol{I_B(x,y)}$.
Figs. 2a and 2b show the original image and the cropped image, respectively.



(a)



(b)

Fig. 2. (a) Original RGB image where the red square is the area to be zoomed. (b) Zoomed cropped image.

2.2. Pre-Processing

Pre-processing aims to improve the quality of the image obtained in order increase the efficiency of segmentation algorithms. This process involves the following image processing blocks: Gamma correction on HSV (Hue, Saturation and Value) and Gamma correction on YCbCr (Luminance, Chrominance Blue, Chrominance Red).

2.2.1 Gamma correction on HSV

The cropped image expressed in its primary components $I_R(x,y), I_G(x,y)$ e $I_B(x,y)$ is converted to the HSV component $I_V(x,y)$ [15].

Gamma correction is applied with logarithmical effect to this component so as to enhance the contrast between the background (epithelium) and the proliferating cells. This will further increase the efficiency of segmentation algorithms. The conversion applied can be expressed as [17]:

$$I_V(x,y) = max\big(I_R(x,y), I_G(x,y), I_B(x,y)\big) \quad (1)$$

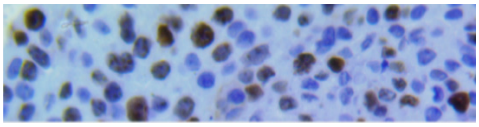
Where $max\big(a_1, a_2, a_3\big)$ function generates the maximum value among a_1, a_2 and a_3.

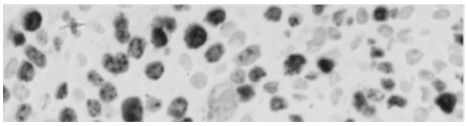
The Gamma correction applied to $I_V(x,y)$can be expressed as [13]:

$$I_{Vg}(x,y) = round\left(\frac{I_V(x,y)}{255}\right)^{\gamma} \quad (2)$$

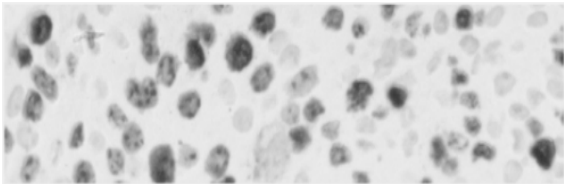
where γ = 0.9 and round(x) return the approximate value to the nearest integer.

Fig. 3a shows image $I(x,y)$ in RGB format; Fig. 3b shows $I_V(x,y)$ after applying (1) to $I(x,y)$; and Fig. 3c shows $I_{Vg}(x,y)$, where non-proliferating or blue cells obtain a colour scheme similar to that of the background, making proliferating cells stand out. Likewise, in Fig. 3c, partially proliferating cells become evident.





(b)



(c)

Fig. 3. Images: (a) $I(x,y)$, (b) $I_V(x,y)$ and (c) $I_{Vg}(x,y)$

2.2.2. Gamma correction on YCbCr

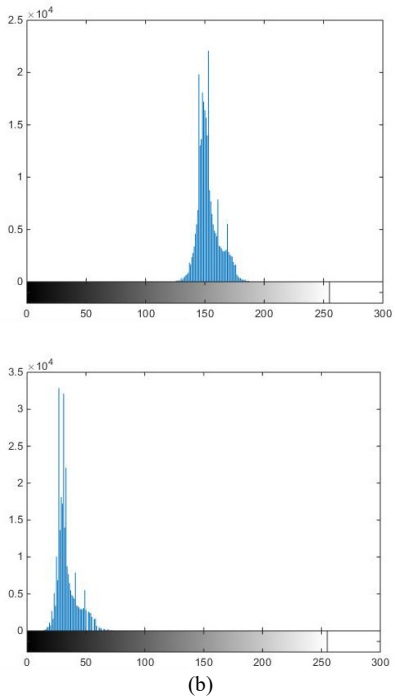
In parallel with the process above, the cropped image expressed in its primary components $I_R(x,y), I_G(x,y)$ and $I_B(x,y)$ is converted to the YCbCr component $I_{Cb}(x,y)$ the conversion can be expressed as [15]:

$$I_{Cb}(x,y) = 128 - 37.797 I_R(x,y) - 74.203\, I_G(x,y) + 112\, I_B(x,y) \quad (3)$$

Gamma correction is applied to this component with γ = 4 (exponential effect). This allows to displace Cb colour schemes from non-proliferating cells. Displacement from the original Cb dimensional space, shown in Fig. 4a. is shown in Fig. 4b. The image resulting from the process is expressed as $I_{Cb2}(x,y)$, as shown in Fig. 5.
The correction applied allows to separate colour schemes and improve future segmentation of the two types of cells.

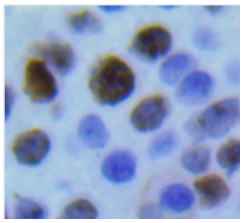
2.3. Segmentation

In this process, objects corresponding to proliferating and non-proliferating cells are segmented. Segmentation is made from the corrected images $I_{Vg}(x,y)$ and $I_{Cb2}(x,y)$ obtained from the original image (Fig. 6).

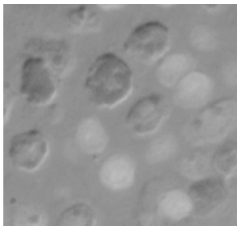


(b)

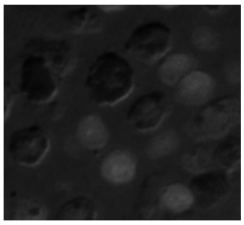
Fig. 4. (a) Original Cb (histogram) dimensional space. (b) Displaced Cb2 (histogram) dimensional space.



(a)



(b)



(c)

Fig. 5. (a) Original image RGB. (b) Image $I_{Cb}(x,y)$. (c) Image $I_{Cb2}(x,y)$ after application of Gamma correction.

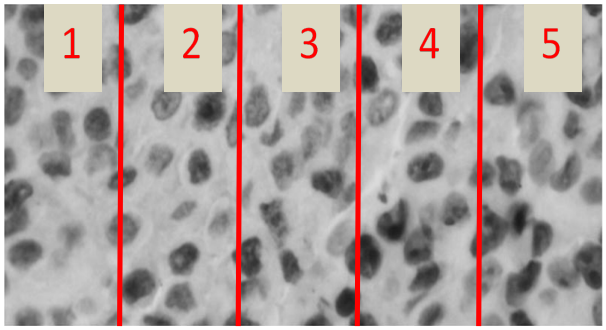


Fig. 6. Image $I_{Vg}(x,y)$ to be subjected to the proliferating cell segmentation process

2.3.1. Segmentation of proliferating cells

To separate the background (cell tissue of no interest for diagnosis) from the proliferating cells, it was decided to use local adaptive thresholds, which result from the probability density function (obtained by histogram) and first order statistical measures (mean and variance). The steps to achieve this process are as follows:

Step 1:

Image $I_{Vg}(x,y)$ is segmented in L' segments. Image segments will have a spatial resolution of M rows by $\frac{N}{L'}$ columns. The segments of image $I_i(x,y)$ can be expressed as:

$$I_i(x,y) = I_{Vg}\left(x, y + i\left(\frac{N}{L'}\right)\right) \tag{4}$$

Where $x = 0,1,\ldots,M-1$; $y = 0,1,\ldots,\frac{N}{L'}-1$ and $i = 0,1,\ldots,L'-1$.

In this case, we have $L' = 5$, because a more even segment is obtained in terms of noise and illumination (Fig.6.).

Step 2:
The histogram of each i segment is determined, which is defined as $h_i(r)$. In this case r represents each colour scheme from 0 to 255 (Fig. 7).

Step 3:
The mean value of each segment i is obtained:

$$\bar{r}_i = \frac{L'}{M.N}\sum_{r=0}^{255} r h_i(r) \tag{5}$$

Step 4:
The standard deviation for each segment i is obtained:

$$\sigma_i = \sqrt{\frac{L'}{MN}\sum_{r=0}^{255}(r-\bar{r}_i)^2\, h_i(r)} \tag{6}$$

Step 5:
The threshold u_i to be used locally for each segment in order to separate cell tissue from the proliferating cell area is the result of the subtraction of the mean and standard deviation:

$$u_i = \bar{r}_i - \sigma_i \tag{7}$$

Step 6:
Thresholding is then applied to the segment:

$$\hat{I}_i(x,y) = \begin{cases} 1\,; & I_i(x,y) \geq u_i \\ 0 & ; \quad I_i(x,y) < u_i \end{cases} \tag{8}$$

Step 7:
Finally, each segment thresholded $\hat{I}_i(x,y)$ is placed in the original order so as to reconstruct an image with spatial resolution of M

rows and N columns, which will be called $I_{pro}(x,y)$ and will contain the proliferating cell area as shown in Fig. 8.

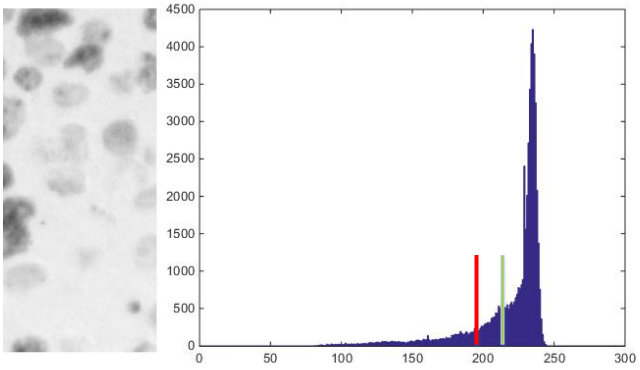


Fig. 7. Example: Segment 1 of $I_{Vg}(x,y)$) and its histogram (in this case, the green bar indicates the mean value of the histogram and the red bar the threshold obtained).

2.3.2. Segmentation by Sobel filter and morphology

After identifying proliferating cells, it is necessary to spot non-proliferating cells in order to find the percentage of cells with proliferative activity in the image under evaluation $I(x,y)$.
For this process, image $I_{Cb2}(x,y)$ obtained in the step above is used. The procedure is detailed below:

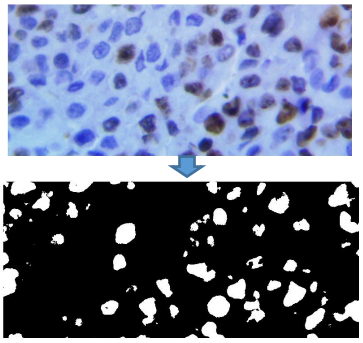


Fig. 8. Original Image and resulting binary image $I_{pro}(x,y)$ after the proliferating cell segmentation process.

Step 1:
A thresholding filter is applied to $I_{Cb2}(x,y)$ in order to segment non-proliferating cells. Thresholding can be expressed as:

$$I_{Cb3}(x,y) = \begin{cases} 0 \quad ; \quad I_{Cb2}(x,y) < 41 \\ I_{Cb2}(x,y) \quad ; \quad other\ way \end{cases} \tag{9}$$

The threshold was obtained from the study and evaluation of image histograms $I_{Cb2}(x,y)$ corresponding to various samples obtained. The result of the process is shown in Fig. 9.

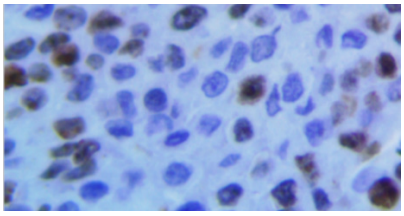
Step 2:
Then, the Sobel filter [17] for edge detection is applied to binary images.
In this process, $I_{Cb3}(x,y)$ is convolved with each of the Sobel masks (horizontal and vertical) which are defined in matrix as [14]:

$$\mathbf{M_v} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{10}$$

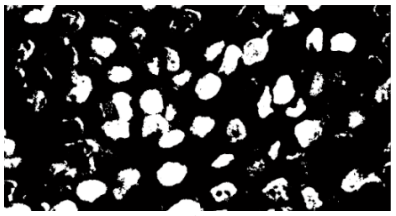
$$\mathbf{M_h} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{11}$$

The resulting image can be expressed as (binary):

$$I_{Sobel}(x,y) = \begin{cases} 1 \ ; \ I'_{Sobel}(x,y) > 6 \\ 0 \quad ; \quad other\ way \end{cases} \tag{12}$$



(a)



(b)

Fig. 9. (a) Original image. (b) Resulting image $I_{Cb3}(x,y)$

Where:

$$I'_{Sobel}(x,y) = |I_{Cb3}(x,y) * M_h(x,y)| + |I_{Cb3}(x,y) * M_v(x,y)| \tag{13}$$

Step 3:
$I_{Sobel}(x,y)$ is subjected to a "closure" morphological operation which can be expressed through the following matrix equation:

$$\mathbf{I}_{clo} = ([(\mathbf{I}_{Sobel} \oplus \mathbf{B}_v) \oplus \mathbf{B}_h] \ominus \mathbf{B}_v) \ominus \mathbf{B}_h \tag{14}$$

Where $\oplus$ indicates the dilation operation and $\ominus$ the erosion operation.
The structural elements used can be expressed as:

$$\mathbf{B_h} = [\mathbf{1\ 1\ 1}] \tag{15}$$

$$\mathbf{B}_v = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \end{bmatrix} \tag{16}$$

In order to reform the cells detected and not to distort the visual result, gap-filling is applied to the matrix cells $\mathbf{I}_{clo}$. This operation can be expressed as [16]:

$$\mathbf{I}_{clo}^{k} = \left(\mathbf{I}_{clo}^{k-1} \oplus \mathbf{D}_p\right) \cap \dot{\mathbf{I}}_{clo}$$
$$k = 1,2,3,\ldots,K0 \tag{17}$$

Where k indicates the number of filling iteration, $\cap$ denotes the intersection operation, and $\dot{\mathbf{I}}_{clo}$ expresses the complement of matrix $\mathbf{I}_{clo}$. Filling is made through $K0$ iterations until condition $\mathbf{I}_{clo}^{k} = \mathbf{I}_{clo}^{k-1}$ is satisfied. The structuring element $\mathbf{D}_p$ used can be expressed as:

$$\mathbf{D}_p = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \end{bmatrix} \tag{18}$$

Finally, the result of this stage is show on Fig.10 and placed in matrix $\mathbf{I}_{Sobelf}$:

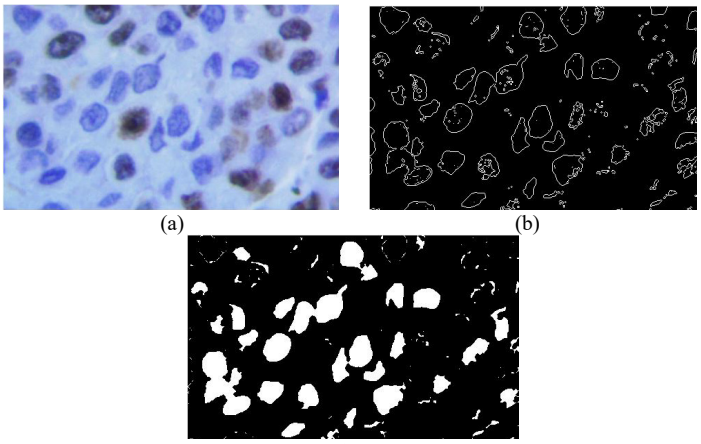
$$\mathbf{I}_{Sobelf} = \mathbf{I}_{clo}^{K0} \tag{19}$$

2.3.3. Segmentation by thresholding on primary components

At this stage, thresholding is applied to the primary RGB components to separate proliferating and non-proliferating cells from the epithelium (background). The thresholds used were obtained from experimental tests performed with images from different tissue samples. The procedure helps remove the area of the epithelium and highlight cells of interest. The steps of the procedure are as follows:

Step 1:
The primary components of the original image **I** are thresholded (Fig 11a):

$$Z_R(x,y) = \begin{cases} 0\ , & I_R(x,y) < 128 \\ 1\ , & I_R(x,y) \geq 128 \ \vee \ I_G(x,y) \geq 128 \end{cases} \tag{20}$$



(a) (b) (c)

Fig. 10. (a) Cropped original image. (b)Resulting Image after applying Sobel Filter. (c)Final Image $I_{Sobel}(x,y)$ after gap-filling.

$$Z_G(x,y) = \begin{cases} 0\ , & I_G(x,y) < 128 \\ 1\ , & other\ way \end{cases} \tag{21}$$

$$Z_B(x,y) = \begin{cases} 0\ , & I_B(x,y) < 128 \\ 1\ , & I_B(x,y) \geq 128 \ \vee \ I_G(x,y) \geq 128 \end{cases} \tag{22}$$

Step 2:
A matrix of zeros $I_{cel}(x,y)$ with spatial resolution $M \times N$ is generated. Then some pixels of $I_{cel}(x,y)$ are changed to 1 from the following condition:

$$I_{cel}(x,y) = \begin{cases} 1\ , & (Z_R(x,y) = 1\ ) \wedge \\ & (\ Z_G(x,y) = 1)\ \wedge \\ & (Z_B(x,y) = 1) \\ 0\ , & other\ way \end{cases} \tag{23}$$

$I_{cel}(x,y)$ finally constitutes a template where cell pixels have value 1 (Fig. 11b).

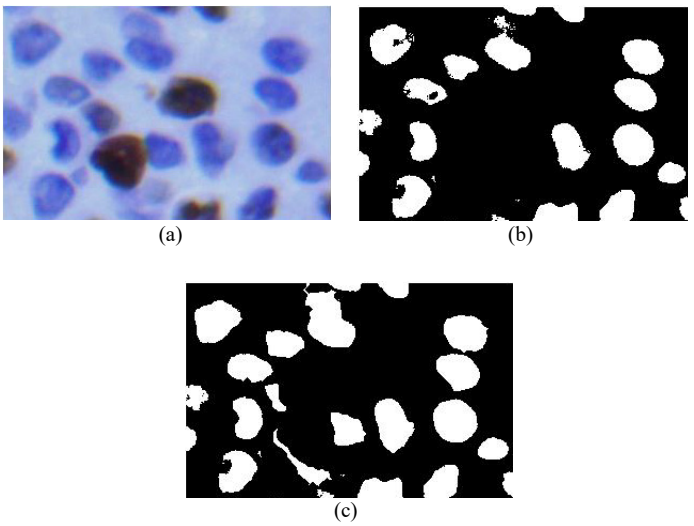
Step 3:
Finally, the matrix of non-proliferating cells (Fig. 11c) is obtained from the following expression:

$$I_{nopro}(x,y) = \begin{cases} 1\ , & (I_{cel}(x,y) = 1\ ) \vee \\ & (\ I_{sobelf}(x,y) = 1\ ) \wedge \\ & (I_{Pro}(x,y) \neq 1) \\ 0\ , & other\ way \end{cases} \tag{24}$$

2.3.4. Segmentation by cell junction

The combination of the proliferating and non-proliferating cell areas results in a binary image containing two types of segmented cells. This image is obtained as follows:

$$I_{total}(x,y) = \begin{cases} 1\ , & \left(I_{pro}(x,y) = 1\right) \vee \\ & \left(\ I_{nopro}(x,y) = 1\right) \\ 0\ , & other\ way \end{cases} \tag{25}$$



(a) (b) (c)

Fig. 11. (a) Cropped original image. (b) Image $I_{cel}(x,y)$. (c) Image $I_{nopro}(x,y)$.

However, one of the problems still present when the cells detected are counted is that some objects are made up of two or more joint or overlapped cells, which results in the algorithm giving a wrong diagnosis. In that sense, as in an acquisition with 40x magnification the average size of a cell is between 200 and 1500 pixels, we can consider that an object of $I_{total}(x,y)$ with a size above 1500 pixels is composed of more than one cell. This will definitely affect the counting, as an object is in principle composed of a single cell. To solve the problem, the following cell separation algorithm is proposed in order to minimise the counting error:

Step 1:
Consider $I_j(x_j, y_j)$ a binary segment of $I_{total}(x,y)$ that contains object j with more than 1500 pixels. In this case, $x_j = 0,1,2 \ldots, m_j - 1$..., and $y_j = 1,2, \ldots, n_j - 1$; i.e. the segment size is $m_j \times n_j$ pixels.

Step 2:
We perform image rotation for different angles θ ($\theta_0 = 0°, \theta_1 = 15°, \theta_2 = 30°, \theta_3 = 60°, \theta_4 = 75°$ and $\theta_5 = 90°$). In this case, the rotation axis passes through the centre of image $I_j(x_j, y_j)$. The range of 15° between rotation angles was chosen based on the best results obtained in various tests of cell separation. Each turn generates an image that can be expressed as $I_j^{\theta}(x'_j, y'_j)$,where θ indicates the rotation angle. The spatial dimensions of each $I_j^{\theta}\left(x'_j, y'_j\right)$ is expressed as m_j^{θ} rows by n_j^{θ}columns.

Step 3:
The horizontal projection $I_j^{\theta}\left(x'_j, y'_j\right)$ is obtained:

$$P_j^{\theta}(y'_j) = \sum_{x'j=0}^{m_j^{\theta}-1} I_j^{\theta}\left(x'_j, y'_j\right) \tag{26}$$

Step 4:
It is important to assess the central area of protection for each $P_j^{\theta}(y'_j)$, which is why the furthest areas of the vector are deleted:

$$\hat{P}_j^{\theta}\left(y'_j\right) = \begin{cases} P_j^{\theta}\left(y'_j\right) , & \left(y'_j > q_0.n_j^{\theta}\right) \vee \\ & \left(y'_j < q_1.n_j^{\theta}\right) \\ 0 , & other\ way \end{cases} \tag{27}$$

Where $q_0 = 0.3$ and $q_1 = 0.7$.

Step 5:
Value $\hat{P}_j^{\theta}(y_0(\theta, j))$is obtained, which is defined as the minimum value of the projection $\hat{P}_j^{\theta}\left(y'_j\right)$. In this case, $y_0(\theta, j)$ is the position where $\hat{P}_j^{\theta}\left(y'_j\right)$ is the minimum value.

Step 6:
The $\mathbf{W}_j$ vector is then obtained:

$$\mathbf{W}_j = \begin{bmatrix} \hat{P}_j^{\theta_0}(y_0(\theta_0, j)) \\ \hat{P}_j^{\theta_1}(y_0(\theta_1, j)) \\ \vdots \\ \hat{P}_j^{\theta_5}(y_0(\theta_5, j)) \end{bmatrix} \tag{28}$$

Step 7:
The minimum value of vector $\mathbf{W}_j$ is obtained, which is defined as $\hat{P}_j^{\theta_{L0}}(y_0(\theta_{L0}, j))$, where θ_{L0} is the angle corresponding to the minimum value of $\mathbf{W}_j$.

Step 8:
Then $Z'_j(x'_j, y'_j)$ is obtained from $I_j^{\theta_{L0}}(x'_j, y'_j)$:

$$Z'_j\left(x'_j, y'_j\right) = \begin{cases} 0 \quad ; \quad y'_j = y_0(\theta_{L0}, j) \\ I_j^{\theta_{L0}}\left(x'_j, y'_j\right) \quad ; \quad other\ way \end{cases} \tag{29}$$

Step 9:
Rotation of $-\theta_{L0}$ over $I_j^{\theta_{L0}}\left(x'_j, y'_j\right)$ is applied to restore the original orientation (although distortion is present due to the interpolation applied to the rotation process). After rotation, image $V_j(x'_j, y'_j)$ is obtained.

Step 10:
Finally, the block of image $V_j(x'_j, y'_j)$ is inserted in $I_{total}(x,y)$ (in the region where $I_j(x_j, y_j)$ was extracted).

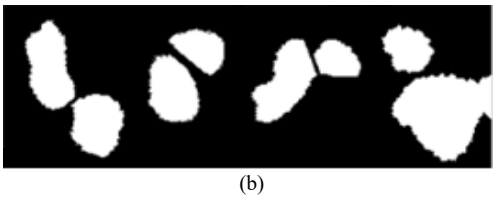
$$I_{total}(xa_j + x'_j, ya_j + y'_j) = V_j(x'_j, y'_j) \tag{30}$$

Where xa_j and ya_j are the coordinates of the first pixel segment extracted $I_j(x_j, y_j)$.
Fig. 12a shows original areas of $I_{total}(x,y)$ containing objects made up of more than one cell. Fig. 12b shows the results obtained after applying the algorithm of cell separation.

2.4. Classification and counting of proliferating and non-proliferating cells

From the total number of cells detected in image $I_{total}(x,y)$, we proceed to classify and count proliferating and non-proliferating cells.

(a)



(b)

Fig. 12. (a) Samples with junction. (b) Samples after the separation process.

Cooperation with pathologists was managed so that the parameters of cell classification were correct and a good diagnosis could be achieved. In this case, doctors requested to classify proliferating cells into three sub groups: strongly proliferating, moderately proliferating and poorly proliferating. The classification criterion is based on the size of the cell area affected by the KI-67 protein.

Image $I_{total}(x, y)$is subjected to a labelling algorithm [16] that will label each cell detected as a k object. In this case, k is the object counter detected in the image. For a total of Q objects, we have $k = 0,1,2, \ldots Q-1$.

2.4.1. Classification of Proliferating and Non-Proliferating Cells

For the classification stage, counters $\mathrm{Total_{pro}} = 0$ (proliferating cell counter) and counter $\mathrm{Total_{npro}} = 0$ (non-proliferating cell counter) are initialised.

$\mathrm{T_{obj}}$ is defined as the total number of objects or cells contained in $\mathrm{I_{total}(x, y)}$. In this context, k is taken to be the object or cell number to be processed. In this case $\mathrm{k} = 0,1, \ldots, \mathrm{T_{obj}} - 1$. We make $\mathrm{k} = 0$ and carry out the following procedure:

Step 1:
A segment $I_k(x^*, y^*)$ of image $I_{total}(x, y)$ is extracted consisting of M_k rows and N_k columns, where $x^* = 0,1, \ldots, M_k - 1$ and $y^* = 0,1, \ldots, N_k - 1$. In this case, $I_k(x^*, y^*)$ is the image segment containing the object k detected in $I_{total}(x, y)$.

Step 2:
A flag f_k is updated through the following expression:

$$f_k = \sum_{x^*=0}^{M_k-1} \sum_{y^*=0}^{N_k-1} I_k(x^*, y^*). I_{pro}\left(x_{0_k} + x^*, y_{0_k} + y^*\right) \tag{31}$$

Where x_{0_k} and y_{0_k} are the coordinates of the first pixel of $I_k(x^*, y^*)$ in $I_{total}(x, y)$.

Step 3:
If $f_k \neq 0$, the cell under evaluation is considered proliferating and the counter $Total_{pro}$ increases by one unit. Otherwise, the counter $Total_{npro}$ increases. Within the requirements of doctors, a sub classification of proliferating cells was made in 3 levels of proliferation: strongly proliferating, moderately proliferating and poorly proliferating.
In this way, when $f_k \neq 0$, steps 4 to 9 are executed.

Step 4:
A temporary matrix $C_k(x^*, y^*)$ consisting of M_k rows and N_k columns is generated and applied:

$$C_k(x^*, y^*) = \begin{cases} 1\,; \; si\; I_{pro}\left(x_{0_k} + x^*, y_{0_k} + y^*\right) \neq 0 \\ 0\,; \qquad other\; way \end{cases} \tag{32}$$

Step 5:
We count each pixel belonging to an area that reacted to KI-67:

$$Apro_k = \sum_{x^*=0}^{M_k-1} \sum_{y^*=0}^{N_k-1} C_k(x^*, y^*) \tag{33}$$

Step 6:
Similarly, for the non-proliferating cell area, a matrix $C'_k(x^*, y^*)$ consisting of $M_k$rows and N_k columns is generated and applied:

$$C'_k(x^*, y^*) = \begin{cases} 1\,, I_{nopro}\left(x_{0_k} + x^*, y_{0_k} + y^*\right) \neq 0 \\ 0\,, \qquad other\; way \end{cases} \tag{34}$$

Step 7:
The pixels of the regions that did not react to the protein KI-67 are counted:

$$Anpro_k = \sum_{x^*=0}^{M_k-1} \sum_{y^*=0}^{N_k-1} C'_k(x^*, y^*) \tag{35}$$

Step 8:
Then the intensity factor of each cell is found:

$$R_k = \frac{Apro_k}{Anpro_k + Apro_k} \tag{36}$$

Step 9:
Flags of proliferation fp_k, fm_k and fd_k are updated:

$$fp_k = \begin{cases} 1\,, & R_k \geq 0.7 \\ 0\,, & other\; way \end{cases} \tag{37}$$

$$fm_k = \begin{cases} 1\,, & R_k > 0.2 \wedge R_k < 0.7 \\ 0\,, & other\; way \end{cases} \tag{38}$$

$$fd_k = \begin{cases} 1\,, & R_k \leq 0.2 \\ 0\,, & other\; way \end{cases} \tag{39}$$

For each k where $f_k \neq 0$ there is only one active flag (fp_k, fm_k and fd_k).
Three counters (initialised to zero), defined as strong, medium and weak, are updated. They are updated as follows: if $fp_k = 1$, the strong counter is incremented by one; if $fm_k = 1$, the medium counter is increased by one; if $fd_k = 1$, the weak counter is increased by one. These counters store the amount of subclassified cells. Note that strong + medium + weak = $Total_{pro}$. This data is granted to the pathologist for evaluation and diagnosis, as shown in Table 1.
Then, k increases by one unit and we go back to step 1. The procedure ends when $k = T_{obj}$.

Step 10:
After evaluating all the cells, we find the percentage of proliferating cells and obtain:

$$\%Diagnosis = \left(\frac{Total_{pro}}{Total_{pro} + Total_{npro}}\right) \times 100 \tag{40}$$

Table 1. Results after test.

Total counted nucleus	Quantity	Percentage (%)
Negative nucleus	94	88
Positive nucleus	13	12
Total	107	100
Positive counted nucleus		
Strong positive nucleus	2	15
Medium positive nucleus	7	54
Weak Positive Nucleus	4	31

3. Results

To prove that the software can be used by specialists, the results obtained by it had to be validated against the results provided by doctors; they were requested to make the same analysis on equal samples. The doctors that carried out the tests are specialised in pathology and work in different institutions in the city of Lima.

Table 2, 3, 4, 5 and 6 show the results obtained in different cases. When comparing the results obtained by the application against the diagnosis given by doctors, it was concluded that, in samples with a clear KI-67-active cell rate (either in a very small or very high percentage in proportion to the total number of cells), doctors using the eyeballing technique obtain results very similar to those of the application. But when samples present some difficulty at the time of counting, it is observed that the results among doctors are poorly correlated with each other and only a few can give a diagnosis similar to that of the application. This is because doctors are subject to different stimuli that influence their performance and they have different counting methods, such as hotspots, random areas, or groups near adjacent hotspots [7], while the software will always proceed in the same way.

In Figure 13 it is shown a section of an evaluated image by both the medical specialist and the developed algorithm. Five cells have been enumerated and marked in a rectangle, named: cell 1, cell 2, cell 3, cell 4 and cell 5.

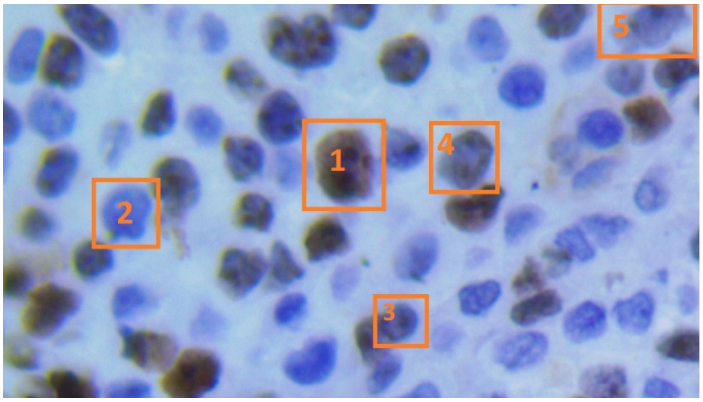


Fig. 13. Cells evaluated by specialist and the proposed algorithm.

For cell 1, both medical specialist and proposed algorithm confirmed that represent a proliferative cell, meaning it was stained in brown color by the IHC process.
For cell 2, again both specialist and algorithm concurred that the cell was non-proliferative, stained with a blue tone by the IHC process. In case of cell 3 and 5, the staining was lower than cell 1, specialist marked the cell as non-proliferative and the algorithm marked as proliferative, is in this scenario in which the subjectivity of human eye turn the proliferation index and can occur divergence in the diagnosis. In case of cell 4, shows partial staining and both specialist and algorithm matched as proliferative.

To demonstrate the error trend, the table below shows the calculus of coincidence between the software and the specialists, and its measuring using the Kappa coefficient.

Defining Kappa coefficient as [18]:

$$\tilde{K} = \frac{\bar{T} - \bar{T}_e}{1 - \bar{T}_e} \tag{41}$$

Define $\bar{T}$ as:

$$\bar{T} = \frac{1}{N_0}\sum_{i=1}^{N_0} \beta_i \tag{42}$$

Where N_0 is the total number of category, in this case two categories $N_0 = 2$: Luminal A, and Luminal B, as explained on [6], Luminal A if diagnosis was <20% and Luminal B if was ≥20%.

β_i is defined as:

$$\beta_i = \frac{1}{\alpha(\alpha - 1)}\left[\left(\sum_{j=1}^{N_0} \alpha_{ij}^2\right) - \alpha\right] \tag{43}$$

Where α is the number of occurrences from a category, α_{ij} is the number of times the specialist chose a diagnosis in the category i and the software chose category j.

Finally:

$$\bar{T}_e = \sum_{j=1}^{N_0} \varepsilon_j^2 \tag{44}$$

Where ε_j is calculated using the following expression:

$$\varepsilon_j = \frac{1}{N_0 \alpha}\left(\sum_{i=1}^{N_0} \alpha_{ij}\right) \tag{45}$$

To calculate precision and recall of specialist vs software, $F - score$ value was calculated [19]:

$$F - score = \left(\frac{(\rho^2 + 1).precision \times recall}{\rho^2.precision \times recall}\right) \tag{46}$$

Where precision and recall were calculated as:

$$precision = \frac{wp}{wp + lp} \tag{47}$$

And:

$$recall = \frac{wp}{wp + ln} \tag{48}$$

Where:
wp: true positive, a cell that specialist said was proliferative and algorithm said was proliferative.
lp: false positive, a cell that specialist said wasn't proliferative and algorithm said it was proliferative.
ln: false negative, a cell that said specialist said was proliferative but algorithm said wasn't proliferative.

$F - score$ is considered balanced so:

$$\rho = 1 \tag{49}$$

Finally, $F - score$ is calculated as:

$$F - score = \frac{2wp}{(2wp + lp + ln)} \tag{50}$$

The average result for F-score of the samples was of 0.8764, showing a high concordance between the software and specialist. The software detects a greater number of cells than the specialist, so the F-score is not perfect, however, these cells not accounted were mostly discarded by specialist errors.
As shown in the Tables 2, 3, 4, 5 and 6 there is a remarkable variation among specialists in the Luminal B diagnosis made by counting. The poor agreement in the diagnosis is due to the difference between stained and non-stained nuclei as rapidly observed by the specialist. Nevertheless, two specialists gave the same result as the software; it should be noted that a proliferation index >20% is needed to share the same diagnosis.
Tables 7, 8 and 9 show the difference among specialists when giving their diagnosis. The Kappa index among specialists is not constant, an average of 0.6863, which demonstrates, once again, the difference generated by the subjective analysis by eyeballing [9][10].

Table 2. Comparison Software vs. Specialist 1

Specialist 1: Kappa $\tilde{K} = 0.689$			
Software \ Specialist 1	Luminal A	Luminal B	TOTAL
Luminal A	7	3	10
Luminal B	0	9	9
TOTAL	7	12	19

Table 3. Comparison Software vs. Specialist 2

Specialist 2: Kappa $\tilde{K} = 0.689$			
Software \ Specialist 2	Luminal A	Luminal B	TOTAL
Luminal A	7	3	10
Luminal B	0	9	9
TOTAL	7	12	19

Table 4. Comparison Software vs. Specialist 3

Specialist 3: Kappa $\tilde{K} = 1$			
Software \ Specialist 3	Luminal A	Luminal B	TOTAL
Luminal A	10	0	10
Luminal B	0	9	9
TOTAL	0	9	19

Table 5. Comparison Software vs. Specialist 4

Specialist 4: Kappa $\tilde{K} = 1$			
Software \ Specialist 4	Luminal A	Luminal B	TOTAL
Luminal A	10	0	10
Luminal B	0	9	9
TOTAL	0	9	19

Table 6. Comparison Software vs. Specialist 5

Specialist 5: Kappa $\tilde{K} = 0.787$			
Software \ Specialist 5	Luminal A	Luminal B	TOTAL
Luminal A	10	0	10
Luminal B	2	7	9
TOTAL	12	7	19

Table 7. Comparison Specialist 1 vs. Specialist 2

Specialist 1 vs Specialist 2: Kappa $\tilde{K} = 0.774$			
Specialist 2 \ Specialist 1	Luminal A	Luminal B	TOTAL
Luminal A	6	1	7
Luminal B	1	11	12
TOTAL	7	12	19

Table 8. Comparison Specialist 2 vs. Specialist 4

Specialist 2 vs Specialist 4: Kappa $\tilde{K} = 0.596$			
Specialist 4 \ Specialist 2	Luminal A	Luminal B	TOTAL
Luminal A	7	0	7
Luminal B	4	8	12
TOTAL	11	8	19

Table 9. Comparison Specialist 1 vs. Specialist 5

Specialist 1 vs Specialist 5: Kappa $\tilde{K} = 0.689$			
Specialist 5 \ Specialist 1	Luminal A	Luminal B	TOTAL
Luminal A	7	3	10
Luminal B	0	9	9
TOTAL	7	12	19

4. Conclusions

In this publication, an algorithm to quantify the proliferative index of the KI-67 marker in histological samples of breast tissue has been developed that proves to give results more reliable than results given by specialist, thanks to image processing techniques, this algorithm can help to provide a better diagnostic to treat breast cancer when applied the result with other biomarkers to diagnose cancer.

A software application was developed. It meets the standards and criteria that pathologists use for the evaluation of the immunohistochemistry method and it is approved by specialists, as they consider that, when testing the software in their work centres, results are reliable.

The iterative algorithm of cell separation is computationally heavy and must be optimised for the comfort of physicians.

In developing the software, in cooperation with medical advisors, new limitations and standards emerged; this is the only way to consolidate a multidisciplinary project.

The quality of the images provided must have some standard sharpness and correct pigmentation to be processed by the application.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] E. Poquioma. “Situación del cáncer de mama en el Perú”. Presentation. Instituto Nacional de Enfermedades Neoplásicas (INEN), 2008.

[2] C. Ramos, D. Venegas. “Análisis de la Situación del Cáncer en el Perú”. Presentation. Dirección General de Epidemiología, Ministerio de Salud del Perú, 2013.

[3] S. Coronato, G. Laguens, O. Spinelli, W. Di Girolamo. “Marcadores tumorales en cancer de mama” MEDICINA (Buenos Aires), 63, 73-82, 2002.

[4] B. Hirata, B. Karina, et al. "Molecular markers for breast cancer: prediction on tumor behavior." Disease markers, 2014. http://dx.doi.org/10.1155/2014/513158

[5] H. Xin et al. “Ki-67 is a valuable prognostic predictor of lymphoma but its utility varies in lymphoma subtypes: evidence from a systematic meta-analysis.” BMC cancer, 14, 153, 2014. http://doi:10.1186/1471-2407-14-153

[6] S. Bustreo, S. Osella-Abate, et al. “Optimal Ki67 cut-off for luminal breast cancer prognostic evaluation: a large case series study with a long-term follow-up” Breast Cancer Res Treat, 157(2), 363–371, 2016. 10.1007/s10549-016-3817-9

[7] L. Fulawka, A. Halon. “Ki-67 evaluation in breast cancer: The daily diagnostic practice” Indian J. Pathol Microbiol, 60(2), 177-184, 2017. http://doi:10.4103/IJPM.IJPM_732_15

[8] Gudlaugsson, E., Skaland, I., Janssen, E. A., Smaaland, R., Shao, Z., Malpica, A., ... & Baak, J. P. (2012). Comparison of the effect of different techniques for measurement of Ki67 proliferation on reproducibility and prognosis prediction accuracy in breast cancer. Histopathology, 61(6), 1134-1144.

[9] Zhong, F., Bi, R., Yu, B., Yang, F., Yang, W., & Shui, R. (2016). A comparison of visual assessment and automated digital image analysis of Ki67 labeling index in breast cancer. PloS one, 11(2), e0150505.

[10] Koopman, T., Buikema, H. J., Hollema, H., de Bock, G. H., & van der Vegt, B. (2018). Digital image analysis of Ki67 proliferation index in breast cancer using virtual dual staining on whole tissue sections: clinical validation and inter-platform agreement. Breast cancer research and treatment, 169(1), 33-42.

[11] V. Anari, P. Mahzouni, R. Amirfattahi “Computer-aided Detection of Proliferative Cells and Mitosis Index in Inmunohistichemically Images of Meningioma” In 6th Iranian Conference on Machine Vision and Image Processing, Isfahan, Iran. IEEE, 2010. 10.1109/IranianMVIP.2010.5941151

[12] Grala, B., Markiewicz, T., Kozłowski, W., Osowski, S., Słodkowska, J., & Papierz, W. (2009). New automated image analysis method for the assessment of Ki-67 labeling index in meningiomas. Folia Histochemica et Cytobiologica, 47(4), 587-592.

[13] M. Bouzid, A. Khalfallah, A. Bouchot, M. Selim, F. Marzani. “Automatic Cell Nuclei Detection: a Protocol to Acquire Multispectral Images and to compare results between colour and multispectral images” in SPIE Photonics West, BIOS, Imaging, Manipulation and Analysis of Biomolecules, Celles and Tissues XI, San Francisco, United States, 2013. doi.10.1117/12.2001980

[14] K. Ahmad, N. Yusoff. “Breast Cancer Types Based on Fine Needle Aspiration Biopsy –Data Using Random Forest Classifier” in I 13th International Conference on Intellient Systems Design and Applications, Bangi, Malaysia. IEEE, 2013. doi.10.1109/ISDA.2013.6920720

[15] R. Gonzalez, R. Woods. Digital Image Processing, 3rd. Ed., Prentice Hall, 2008.

[16] T. Chun-Ming, L. His-Jian. “Binarization of Colour Document Images via Luminance and Saturation Colour Features” IEEE Transactions on Image Processing, 11(4), 2002. 10.1109/TIP.2002.999677

[17] R. Gonzales, R. Woods, E. Eddins. Digital Imaging Processing Using Matlab, 2nd Ed., Mcgraw Hill, 2010.

[18] J.L Fleiss. “Measuring Nominal Scale Agreement Among Many Raters” Psycological Bulletin, 76(5): 378-382, 1971. http://dx.doi.org/10.1037/h0031619

[19] M. Sokolova, N. Japkowicz, S. Szpakowicz. “Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation” In AI 2006: Advances in Artificial Intelligence, Lecture Notes in Computer Science (vol. 4304), 2006. DOI: 10.1007/11941439_114

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 212-221 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Alternative Financing Model for Smart Cities Initiatives in Indonesia

Robertus Nugroho Perwiro Atmojo[1,*], Tota Pirdo Kasih[2], Yakob Utama Chandra[2]

[1]*Department of Information Systems, School of Information Systems, Bina Nusantara University, Jakarta, Indonesia*

[2]*Department of Industrial Engineering, Faculty of Engineering, Bina Nusantara University, Jakarta Code, Indonesia*

ARTICLE INFO

Article history:
Received: 24 October, 2019
Accepted: 25 November, 2019
Online: 22 January, 2020

Keywords:
Equity crowdfunding
Alternative financing
Indonesia Retail e-Commerce

ABSTRACT

Currently Indonesia is working actively in the implementation of the Smart Cities initiative program. However, to be able to apply the concept, there are three main elements that must be prepared as the basic foundation. These elements are structure, infrastructure, and superstructure. In this study we focused on the problem of the high cost of funding needed to finance these three elements. Therefore, we are of the opinion that the government must create new investment opportunities outside the stock exchange whose impacts can also be directly used for financing programs. In this study, we conducted an investigation of the interest of urban people to invest in online channels. The aim is to determine the decision-making behavior of the Indonesian people in non-consumptive capital expenditure activities when the government creates and publishes financing instruments on the market. The hope is to be able to create a national development financing strategy that is in accordance with the characteristics of the Indonesian people and right on target.

1. Introduction

During 2017 - until the present day, Indonesians' through several collaborative ministries (central government), local governments, the private sector, and the community have a high enthusiasm to improve and build a better nation. Led by the Ministry of Communication and Information, Ministry of Home Affairs, Presidential Staff Office, Ministry of Public Works and Public Housing, along with several other vital agencies. Currently, the nation is focusing on realizing a national development program in various aspects concerning the Smart Cities concept. The Smart Cities understood as a concept in which a region can manage its resources and support the living needs of its ecosystem independently. According to [1], [2], The Smart Cities concept includes six main aspects, namely Smart Governance, Smart Branding, Smart Economy, Smart Living, Smart Society, and Smart Environment. However, before being able to realize the Smart Cities program, it turns out that there are basic needs that must be met before being able to apply the concept.

Based on information obtained from the testimony of one of the compilers of the Guidance Book of Indonesia Smart City Master plan 2017, Ms. Dwi Elfrida (Person in Charge Indonesia Smart City Initiatives from Ministry of Information and Telecommunications). The implementation of the Smart City concept is analogous to a family that must already have a place to live before they can have a car for personal use. In other words, there are three main elements which are the pillars for implementing the Smart Cities concept in Indonesia, namely structure, infrastructure, and superstructure [2]. The structural elements include the development of human capital which is the agent and beneficiary of the Smart Cities concept. Within the structural elements are also discussed about budget resources and local government resources. The second element is infrastructure; infrastructure development includes three components, namely physical, digital and social infrastructure for the advantage of the community. The last element is the superstructure, in this section component such as regional policies and regulations, institutions, and procedures for implementing the Smart Cities concept must be accommodated.

The construction of three main elements are constrained by funding issues

Without having three adequate main elements, it is challenging for a region or even a country to implement the Smart Cities concept [3]. To be able to succeed in applying the concept, at least

* Robertus Nugroho Perwiro Atmojo, Email: robertus.atmojo@gmail.com

the people of a region must have a level of compliance and discipline towards proper legal regulations. Then to support innovation and economic growth, the potential must be supported by robust infrastructure and facilities. Finally, to be able to create a stable civilization on society, legal planners and implementers with integrity and reliability are needed. Unfortunately, in Indonesia, we are still struggling to resolve these crucial issues. The geographical conditions in the form of archipelagic countries and various types of local wisdom, make the development process of Indonesia have many challenges. Also, we suspect, one of the biggest challenges in the process is about infrastructure problems.

Infrastructure development is one of the essential elements that must be carried out by the state to make the nations' wheel of life continue to keep pace with the times. Without the availability of access to highways, railroads, seaports, airports, warehouses, optical fiber, radio towers, power plants, clean water treatment plants, and various other physical and digital infrastructures. So it is impossible for a country to continue to survive and develop in an era of increasingly intense global competition [4]. Unfortunately, infrastructure development is not as easy and cheap as imaginable. In Indonesia, to build the Trans Java toll road project along the 522 Kilometers, it is estimated that it requires an investment cost of IDR 17 Trillion (US$ 1.5 Billion), and even then only for the land acquisition budget [5]. Building nine dams in East Nusa Tenggara, West Nusa Tenggara, Banten, Central Java, Batam, and South Sulawesi, the government estimates that the project may require a national budget of IDR 3.85 Trillion (US$ 300 Million) [6]. Also, for the construction of the Trans Kalimantan railroad network infrastructure for 2400 Kilometers, it is estimated that the development requires an investment cost of IDR 22 Trillion (US$ 2 Billion) [7]. Regarding the example of the Trans Kalimantan infrastructure development project, it is regrettable that the project was one of fourteen lists of National Strategic Projects that had to be canceled by President Joko Widodo in 2018 [8].

Seek for alternative national development funding

The cancellation of the implementation of fourteen National Strategic Projects by President Joko Widodo made sense, considering the required investment costs were estimated to reach IDR 264 Trillion (US$ 20 Billion) [9]. Whereas at the same time, the nation still had to share its focus on the development of structural and superstructure elements. The Indonesian government's efforts to continue development also do not just stop. Issuance of investment instruments such as Saving Bonds Retail series 004 has been issued by the Ministry of Finance in 2018 to obtain a national financing capacity of IDR 1 Trillion (± US $ 100 Million) with coupon rates of 8.05% per annum [10]. In 2019, through the Ministry of Finance, the Indonesian government again issued Saving Bonds Retail series 005 for a national financing capacity of IDR 5 Trillion (± US$ 500 Million) with coupon rates of 8.15% per annum [11]. Government initiatives in issuing Saving Bonds indicate that for continuing national development, there are still many alternative financing sources needed. The good news from the issuance of the Retail Saving Bonds is quite exciting regarding the interest of the Indonesian middle class to contribute their capital toward national development. We assume that the community intentions formed are supported by trust building mechanisms that have been successfully campaigned by the government. Furthermore, the availability of a variety of instruments and channels for investment access both online and offline also helped facilitate community action to invest. Therefore, we suspect that now is the right moment for the government to conduct national revenue re-engineering. Specially to experiment in creating innovative instruments and investment access channels. In this study, we attempt to investigate the interest of urban people to invest in online channels. The aim is to determine the decision-making behavior of Indonesian people in non-consumptive capital expenditure activities. The results obtained from this study are expected to create social development financing opportunities that are following the characteristics of the Indonesian people.

2. Literature Review

A lot of new types of business offering profit sharing and business capital cooperation by financially based technology startups became our source of inspiration in this study. In Indonesia, the Go-Jek transportation ride-sharing startup that has adapted the Uber concept has successfully transformed into one of the financial technology startups that are quite important for the empowerment of local communities. Tokopedia and Bukalapak have fostered an entrepreneurial spirit for workers and retirees to continue developing their businesses. The Bareksa Investment Marketplace allows people to develop their capital not in conventional bank deposits, but on mutual fund exchanges. Bank Negara Indonesia also creates online securities products (with a zaisan and e-smart platform) that enable students to learn stock investments on the market with only less than US$ 10 deposits. More interestingly, in America, there is a crowdfunding-based online funding platform known as the Kickstarter. Where based on the company's claims, currently a total of successful projects funded have reached more than 250,000 projects — supported by around 15.5 million philanthropists, with funding turnover reaching US$ 4 billion.

2.1. Indonesian version of Kickstarter

Technological startups of crowdfunding-based business such as Kickstarter also exist in Indonesia. We know KitaBisa and Mekar as two examples of crowdfunding startups that collect enough philanthropists for business development, even for social financing. For KitaBisa, they claim that the donations have been collected reached IDR 512 Billion (US$ 45 Million). Indonesia also has iGrow, a crowd-investing based marketplace platform that brings farmers (or breeders), owners of agricultural land (or livestock), and investors synchronically. From the business model, iGrow claims to have empowered as many as 2200 farmers to manage more than 1,000 hectares of land, with yields of more than 500 tons only for peanut commodities (not including others) in Indonesia. Today's online channel-based crowdfunding and crowd-investment trends inspire us to test the intentions of the interest of the Indonesian people in making decisions to fund investment and charitable/social activities. To support the investigation, we chose retail-based e-Commerce channels as a testable activity.

2.2. E-Commerce channel

The e-Commerce channel is a dynamic, complex, and mysterious business ecosystem. There are many factors could

make a potential consumer just do online windows shopping or make a real decision when they visit online equity crowdfunding platform or an e-Marketplace. The factors can be influenced by technology or can be influenced by social factors. An excellent platform appearance, content information conformity, and ease of use are some fundamental aspects of technology that could attract customers to make purchases [12]–[14]. On the other hand, the ability of a retailer to provide protection for shopping and the quality of products offered is also an important aspect for motivating customer to purchase or to invest electronically [15]. Not only there, we assume there is also another factor that causes potential customers deciding to make a purchase, that factor is the quality of service [16], [17]. In our perspectives the world of retail e-Commerce, the good and poor quality of retailer services can be judged in two ways. The first aspect is the retailer's experience in fulfilling customer orders. The success of the retailers in completing the transaction process and meeting the needs of customer orders in online business can affect the reputation of the business entities both in the realm of online and conventional sphere. Because in previous study by [18], mentioned that the success of a retailer in delivering a satisfying shopping experience will build a sustainable online trust and it will impact on the seller electronic image. In the other study by [19] also proves that a good electronic image (e-Image) even can trigger consumer interest to pay for a product at much higher price than the suggested retail price (MSRP). The second aspect, we speculate that a retailer's ability to provide customer support/help is one of the keys to success that leads consumers to actualize their purchases. The previous research by [20], It is proof that the product recommendation provided by the retailer can make a potential consumer change product purchase decision. They could switch the purchase intentions of pre-planned products, into buying products recommended by the seller. Therefore, we believe that the presence of retailers in the online channel is also required. Through the electronic Word of Mouth [21], [22] and instant messaging technology for real-time online support assistance [23], [24]. Then, we think that Social Presence in the realm of online transactions is also important to be held.

2.3. Vendor reputation

Vavilis, Petković, & Zannone [25] defined reputation as a measure of online trust. The mechanism of reputation ranking of online business by consumers done in several approaches. As with the eBay auction site or Indonesian e-marketplaces like Tokopedia and Bukalapak, reputation ranking systems use star-scale systems (1 [bad] - 5 [good]) as a sign of the level of business reputation. In the Kaskus community forum, the user reputation rating system uses the colored beam scale (red: bad - Green: best) as the indicator. In a previous study by Gregg & Walczak [19], proving that high reputation ranking value and also a selection of the right business username can affect the level of customer trust. Where ultimately the level of trust formed correlates with the consumer's purchase decision [26]. Chen & Chou [27] defined the vendor's reputation as a reflection of the seller's ability to deliver the best service in the eyes of consumers.

The definition of vendor reputation in this research defined as vendor capability that manifests in recognition and public awards for producing the consistenly good performance. In other words, the positive digital footprint that vendors have because of good business history. Later it will be used as a measurement tool for prospective investors to establish business cooperation with them in the future (see Fig. 1). The measurement indicators of the supplier's reputation consist of:

(i) *The vendor's existence level*. The existence of suppliers in digital and physical channels must be easily recognizable and known to the general public. (ii) *The recognition/commendation level*. The amount of acknowledgment or award earned by a vendor from a community or third party appraisal institution. (iii) *Vendor ratings*. The amount of customer testimonials about the vendor on every products/services sold in internet.

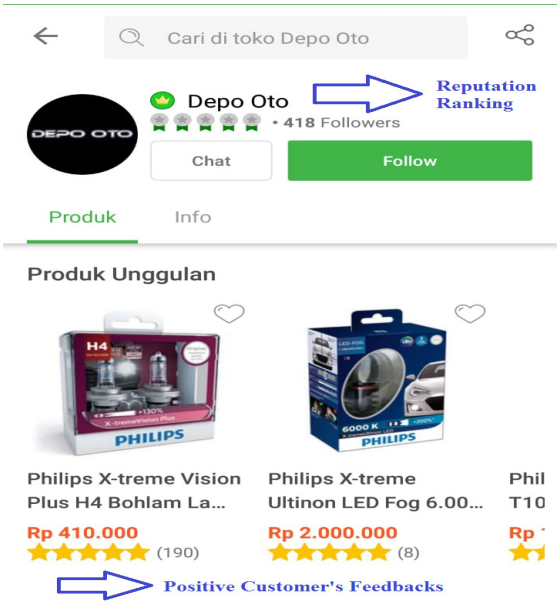


Figure 1. The example of vendor good reputation rating and positive customers' feedback (Depo Oto in Tokopedia.com)

2.4. Social presence

Social presence is a vendor/retailer manifestation in a business transaction mechanism. Social presence is difficult to create in online channels, especially if only rely on graphics or images [28]. However, as time passes and dynamic technological updates trigger the emergence of virtual reality concepts making social presence possible in online channels. The website is one of the key aspects in e-Commerce transaction. According to [22] social representation of online business players, psychologically possible presence by consumers through a particular medium, where the website or mobile site is the intended intermediary. Furthermore, to strengthen that perspective, it is appropriate that every online business person has a website or mobile media that is independent and not bound by third party intervention. Because, through independent ownership, the business has an unlimited discretion in serving consumers and perform regulation.

Further, ownership is personal as well as creates consumer perceptions of credibility, empathy, integrity, and benevolence (good impression good behavior). Ogonowski et al. [29] defined social presence as a perception of the personality, hospitality, and elements of the human sensitivity created in a website. In this

research the social presence is defined as a social approach given by the vendor to their consumer to reduce the service gap that occurs in the digital environment. In this case, the social presence measurement indicator divided into: (i) *Attention level*. Attention provided by the vendor in interacting and communicating with consumers online. (ii) *The vendor's ability level*. The ability of seller in delivering an atmosphere of online shopping experience that resembles the state of shopping traditionally to consumers. (iii) *Vendor politeness*. The ability of seller to express manners and ethics to their customers (see Fig. 2).

Figure. 2. The response from ZOTAC customer service agent that shows politeness and responsible business ethics for customer complaints in Newegg platform.

2.5. Online trust

Gefen [30] stated that consumer confidence in vendors is the basic necessity to obtain second, third, and onward purchase intentions. Alternatively, in other words, consumer loyalty is a manifesto of business success in building trustworthiness in the presence of consumers. Turilli et al. [18] also stated that trusts can not only occur in the real world (offline channels), but it can also form under online conditions. Therefore, the ability of vendors to build trust in the reality will reflect its integrity in digital channels and apply vice versa. Gefen [31] defined trusts in the context of e-Business and e-Commerce as the ability of consumers to take risks of uncertainty situations in transactions that can not be controlled by itself but still take action on certain expectations. According to [28], a definition of trust in the context of e-Business and e-Commerce is the ability of consumers to accept vulnerability owned by online vendors based on positive expectations and certain actions. Furthermore, Hwang & Lee [32] defined online trust as a consumer's ability to ignore uncertainty avoidance by using a reference to cultural norms prevailing in certain society as a guide to reducing risk of worry. In this case, allegedly consumers have a certain belief that things that are harmful to others are not appropriate to be done by someone (the trustee) because if it happened then the parties will receive certain social sanctions.

The ability of vendors to provide transparent, detailed, and accurate information is a trigger for consumer confidence online. Because in remote condition, consumers can not test the product directly, so it has become a duty for a seller to provide the online shopping experience that is not too much different from the traditional shopping experience (Fig. 3). This study defined online trust as a risk-taking mechanism conducted by consumers to depends on the performance capabilities of a vendor in conducting business transactions based on digital channels. And we agreed to have online trust consist of four dimensions, such as integrity, benevolence, ability, and predictability. The dimension of integrity measurement in this study consist of:

(i) *The level of confidence*. Consumers believe their familiar vendor will always be honest and will not cheat them or other customers. (ii) *Reliability level*. Consumers understand their online subscribed vendor will always have a reliable performance and full sense of responsibility for all actions that will accomplish. (iii) *The level of openness*. Online vendors must have transparency or openness in communicating with their customers so that consumers do not have to worry about cheating. Benevolence dimension (good behavior) in this study defined as the ability of the vendor to be able to act well and considerately in every action to the consumer. The measurement indicators of benevolences consist of (iv) *Confidence level*. Consumers believe that online subscribers always prioritize the interests of consumers compared with personal interests. (v) *Confidence level*. Consumers believe that their online subscribers will always act with consideration before performing their actions.

The ability dimension in this study defined as the level of competence possessed by the vendor in providing the best service for its customers. The measurement indicators of the ability dimension consist of: (vi) *The level of vendor competence*. Vendors must be able to demonstrate the competencies expected in providing shopping assistance to their consumers. (vii) *The rate of response speed*. It means, as a time of service that vendors can provide when consumers need help in the shopping process. (viii) *The vendor's troubleshooting competence*. A seller capability required in providing troubleshooting solutions in the event of a dispute with the consumer.

The dimension of predictability in this study defined as the ability of vendors to show performance patterns that can be predicted by consumers through a series of actions that follow the work procedures. The measurement of predictability indicators are as follows: (ix) *The rate of response speed*. The responsivity of vendors when answering questions and providing relevant advice needed by consumers. (x) *Expected level*. Consumers expect to always get an enjoyable online shopping experience from online vendors.

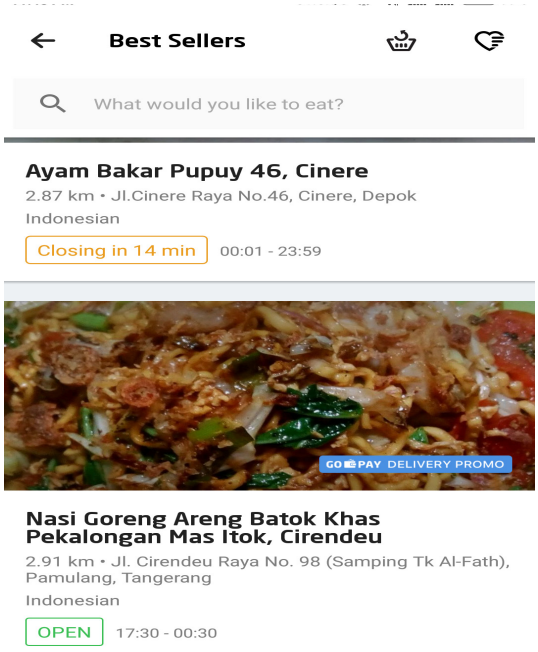


Figure 3. Nasi Goreng "Mas Itok" food vendor that are most preferred by Go-Food (Go-Jek) customers

2.6. Customer investment (purchase) intention

In this study, customer investment (purchase) intention is defined as the motivation of consumers to make investment decision towards equity offered (like in startup business cooperation opportunities). In other words, customer investment (purchase) intention in this research meant as a behavioral-based intention owned by a consumer to visit and make the realization of online purchases (or investment) from the desired business vendor [33]. The intention and desire are the last stages of the various considerations that consumers have before making a final purchase decision. Referring to Hong & Cho [34] measurement indicators of customer investment intention consist of (i) customer will return to visit the online store/vendor. (ii) Customer will make a purchase/investment from the online store/vendor within a period of two months. (iii) While new products are released in the market, then customer will prioritize purchases/investing from certain online stores/vendors. And (iv) for further purchases (or investment), customer will continue to visit and make actual decision from certain online vendor/retailer store.

3. Method

To support the analytic process in this research. We used convenience sampling as a method of data collection. The non-probabilistic sampling approach in this research is used to overcome the problem of dynamic and infinitive samples. For information, the sample used focuses on potential consumers who had shopping experience at online retail stores incorporated into the Indonesian e-Commerce Association (IdEA). As for some examples of these vendors are: included in the category of e-Marketplace (Tokopedia, Lazada, Bukalapak, BliBli) or providing marketplace platform (Go-Jek, and Grab) and also listed in the category of online retail stores (Traveloka, JD.id, Bhinneka).

The design of this study began with a scenario that questioning respondents intention to invest in online startups initiated by their familiar vendors. When their well-known vendors like Depo Oto (Fig. 1), Zotac (Fig. 2), or Fried Rice "Mas Itok," (**Fig. 3**) try to diversify their business in other fields. For example, Depo Oto, which is involved in the business of retailing automotive spare parts, will expand its business into agriculture development. Alternatively, Zotac, which is engaged in the Graphical Processing Unit, expands in the textile sector. And, if fried rice "Mas Itok" who started the business as a food seller innovating to build a semi-real estate property business. With consideration of all forms of administration and business legality guaranteed by Indonesian Government (like Otoritas Jasa Keuangan or The Ministry of Finance). Then we asked the respondents by several questions like:

"*Are the respondents interested on investing their capital in these startup companies?"* Considering they have had a business relationship (experienced) in the past. "*Can social capital that has been formed by both parties be developed to continue business relations and kinship between the two?".*

Some of these questions will then converted and tested through the relationship of several latent variables on the structural model. For the data collection technique, we used one shot cross-sectional and close-ended questionnaire based on a Likert scale with a range of values 1-5 (low-high). For data transformation process from qualitative scale to the quantitative scale used Method of Successive Interval. The goal of scale transformation helps to get a more weighted, stable, and objective assessment results. After the data collection process finished, the next phase is the process of data cleansing. The data cleansing process focused on using Mahalanobis Distance method to control outliers. The process of outliers handling is used to avoid the results of research from the use of biased data. In the last stages of data analysis, we used the method of Structural Equation Modeling based on Partial Least Squares (SEM-PLS). The purpose of using SEM-PLS in this research is to adjust the exploratory nature of research, where it requires a high accuracy measurement result by utilizing the nature of multiple regression analysis.

Refer to Fig. 4, the proposed hypotheses in this study consist of:

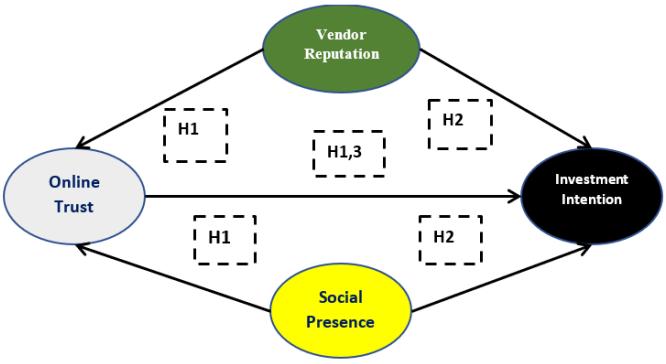


Figure 4. Relationship of structural model proposed in this research

H1: Vendor Reputation and Social Presence are positively correlated and have a significant direct effect on the establishment of Online Trust.

H2: Vendor Reputation and Social Presence are positively correlated and have a significant direct effect towards customer Investment Intention.

H3: Online Trust is positively correlated and has a significant direct effect towards customer Investment Intention.

4. Results and Suggestion

Out of the 350 distributed questionnaires results obtained as many as 323 respondents' answers are valid. The sample respondents are undergraduate students from a university in Jakarta majoring in information systems program. The average age of respondents is in the range 17-28 years. A total of 173 respondents are female, and the rest are male respondents. Based on the sample used, it assumed that participants are potential customers who have experience in the use of computers and smartphones. Moreover, it assumed they have a good understanding of internet usage [35]. Average monthly online spending from respondents ranged from IDR 200,000 - IDR 4,000,000 (US$ 15 – US$ 350). As for the data obtained, the most popular goods purchased by respondents when shopping online is a product of fashion and electronics. Fig. 5 shows the results of the first stage inner model testing using Smart-PLS software version 3. The first step of structural model testing can be seen in Fig. 5.

According to [36], we can say that the Online Trust moderately explains the creation of 40.5% variance in customer Investment Intention. Where, Vendor Reputation and Social Presence also have moderate explanation in Online Trust with the value of 42%. According to the path coefficient value, Online Trust has the strongest correlation towards customer Investment Intention with .457 value. Followed by Vendor Reputation → Online Trust with .366 value. Then Vendor Reputation → customer Investment Intention with .353 and Social Presence → Online Trust with .326 value. The weakest correlation coefficient indicated by Social Presence towards customer Investment Intention with -0.132 value.

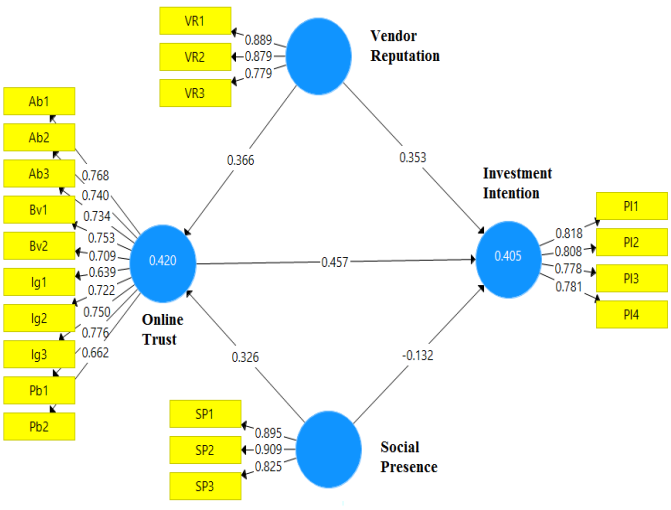


Figure 5. Inner model testing on structural model

In Table 1 it can be seen that all research variables have values of Cronbach α > 0.7 and rho_A > 0.4. In other words, we can say that all variables have met the required reliability standard in the measurement. In Table 1 it can also be seen that the value of Composite Reliability has exceeded 0.6 and the Average Variance Extracted has surpassed the value of 0.5. Thus, we consider all the measurement variables used to meet the criteria of structural reliability model.

Table 1. The Measurement of Reflective Outer Model

Latent Var.	Cronbach α	rho_A	Composite Realibility	Average Variance Extracted (AVE)
Online Trust	.90	.902	.918	.528
Investment Intention	.81	.817	.874	.635
Social Presence	.849	.849	.909	.769
Vendor Reputation	.807	.818	.887	.723

In Table 2 it can be noted that based on the calculation of Fornell-Larcker Criterion method all variables have met the required validity conditions. It indicated by the value of the intended variable (marked with the * sign) has exceeded the correlation value of other variables that reside in each column. For example, in the first column, the Online Trust variable has a diagonal value of .726.

That value exceeds the value of the correlation of other variables: Investment Intention (.593), Social Presence (.603), and Vendor Reputation (.612). In the second column also shows that the Investment Intention diagonal value has (.797) exceeded the Social Presence variable value (.410) and Vendor Reputation (.533). Thus, it can be said that based on the measurement of discriminant validity, then all research variables used have met the requirements of structural validity.

Table 2. Discriminant Validity Measurement

Latent Var.	Online Trust	Investment Intention	Social Presence	Vendor Reputation
Online Trust	**.726***	-	-	-
Investment Intention	.593	**.797***	-	-
Social Presence	.603	.410	**.877***	-
Vendor Reputation	.612	.533	.756	**.85***

4.1. Analysis after bootstrapping procedure

To perform hypothesis testing in SEM-PLS required bootstrapping procedure. The goal is to get more accurate measurements by combining core samples with sub-samples data. In this measurement, we use 5000 data as a sub-sample with two-tailed test mode, no-sign changes, complete bootstrapping, 2GB of memory usage, and 323 data as core samples.

Fig. 6 shows the results of inner model testing after going through the bootstrapping process. Table 3 shows the structural inner model test results after going through the bootstrapping procedure.

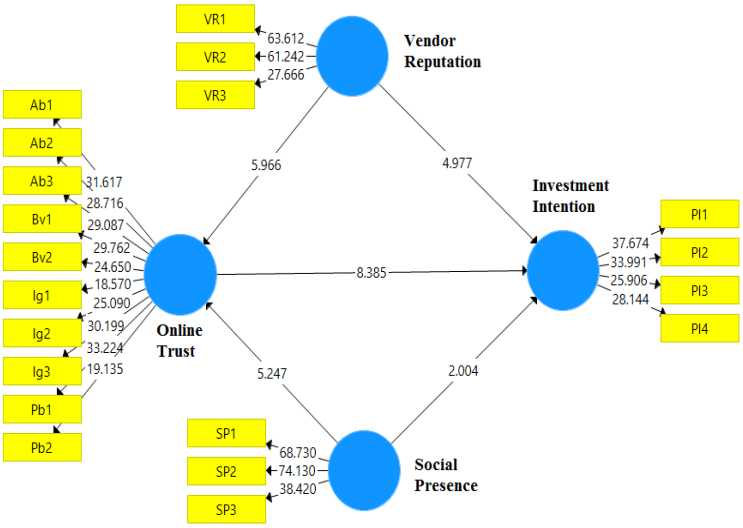


Figure 6. Inner model structure after bootstrapping

Table 3. Path Coeff. Values After Bootstrapping

Latent Var.	Original Sample	Sample Mean	StD	T-Stat	p-Values
Online Trust → Invest. Intention	.457	.459	.054	**8.385**	**.000**
Social Presence → Online Trust	.326	.327	.062	**5.247**	**.000**
Social Presence → Invest. Intention	-.132	-.132	.066	**2.004**	**.045**
Vendor Reputation → Online Trust	.366	.367	.061	**5.966**	**.000**
Vendor Reputation → Invest. Intention	.353	.353	.071	**4.977**	**.000**

The result of path coefficient measurement shown in Table 3 shows that in general, the correlation among the hypothesized variables has a good relationship. It is proofed by the correlation between Online Trust → Investment Intention through the value of T-Stat 8.385 > 1.96 (p-Val < .05). Furthermore, the value of Social Presence correlation → Online Trust has value T-Stat 5.247 > 1.96 (p-Val < .05). The correlation value of Social Presence → Investment Intention has T-Stat 2.004 > 1.96 (p-Val < .05). Correlation value of Vendor Reputation → Online Trust 5.966 > 1.96 (p-Val < .05). Moreover, the correlation between Vendor Reputation → Investment Intention has a T-Stat value of 4,997 > 1.96 (p-Val < .05). Where according to Wong [37] with a significance level of 5% then a good correlation should meet the requirements of T-Stat > 1.96 and p-Val < .05.

4.2. Hypotheses testing

In Table 4 we described the final result of structural model analysis. We use total effects as a measure of hypotheses testing.

Based on the result of total effects testing shown in Table 4. It can be concluded that:

H1: Vendor Reputation and Social Presence are positively correlated and have a significant direct effect on the establishment of Online Trust. (**H1 accepted**, p-Val < .05 & p-Val < .01).

Table 4. The Total Effects

Latent Var.	Original Sample	Sample Mean	StD	T-Stat	p-Values
Online Trust → Invest. Intention	.457	.459	.054	**8.385**	**.000****
Social Presence → Online Trust	.326	.327	.062	**5.247**	**.000****
Social Presence → Invest. Intention	.017	.018	.069	**.251**	**.801**
Vendor Reputation → Online Trust	.366	.367	.061	**5.966**	**.000****
Vendor Reputation → Invest. Intention	.52	.522	.070	**7.428**	**.000****

H2: Vendor Reputation and Social Presence are positively correlated and have a significant direct effect towards customer Investment Intention. (**H2 rejected**, Social Presence → Investment Intention has p-Val > .05).

H3: Online Trust is positively correlated and has a significant direct effect towards customer Investment Intention. (**H3 accepted**, p-Val < .05 & p-Val < .01).

From our measurement results, we found that Vendor Reputation and Social Presence variables have a strong correlation with the Online Trust it shown with the T-Stat > 1.96 and p-Values < .05 for both variables. Vendor Reputation and Social Presence together also have a significant influence on the establishment of Online Trust, it is shown with the value of p-Values < .05. Furthermore, the Online Trust Variables also have a strong correlation with customer Investment Intention variables; it has shown with the value of T-Stat > 1.96 and p-Values < .05. Similarly, the results of total effects measurement, Online Trust variable has a significant influence on the formation of customer Investment Intention with p-Values < .05.

What is interesting in this study is that we found the fact that although the Social Presence has a good correlation with customer Investment Intention (T-Stat [2,004] > 1.96 & p-Val [.045] < .05). However, the Social Presence variable does not have a significant effect on the formation of customer Investment Intention. It is shown with p-Val of total effects that exceed the 5% significance value (.801 > .05). We believe, the influence of Social Presence is not significant to the formation of customer Investment Intention caused by the sample respondents used is the people with a bachelor degree.

In other words, respondents with a decent level of education correlate with their level of understanding about the product and the use of a good e-Commerce retail platform as well. Therefore, allegedly that when they come to investing their capital towards products/services/equities online, respondents from academic societies prioritize vendor reputation compared to the sense of representation of the retailer. So we conclude that as long as the retailer can present a qualified product/service (worth to funded) and followed by a good reputation, then the consumer (especially from academic circles) will still make an investment, regardless of social presence of the retailer.

4.3. Opportunity for Indonesian government

The results of our investigation show that for university students, with a regular budget of IDR 200,000 - IDR 4,000,000 (US$ 15 - US $ 350) per month. The interest in funding vendors with excellent products and reputations has positive results. In other words, this is an opportunity for the government to take advantage of current possibilities regarding National Strategic Projects funding, especially infrastructure. We see opportunities for "government logos" as shown in **Fig. 7** will be an added value for the government when conducting investment campaigns for alternative funding. According to [38] Logos known as trust marks can be useful as accelerators for the creation of trust building mechanism on digital channels. Because with the existence of trust marks, we can say that the product/service/equity issued in the market has passed specific complex procedures and can also be used as an illustration of the reputation vendor.

Figure 7. Trust-Marks from the government

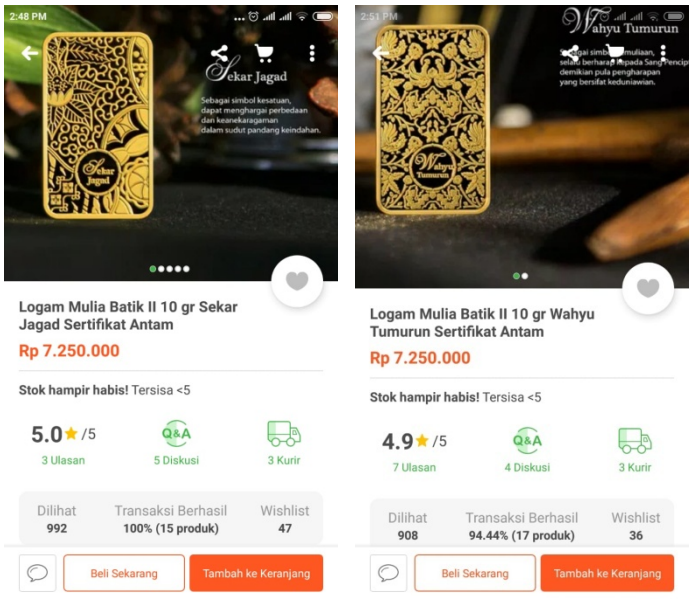


Fig. 8. Products from very good reputable vendor combined with arts are more expensive than regular products. Source: Toko Agung Cikini (Tokopedia's gold/jewelry retailer)

Government campaigns regarding investment programs for the community can be put into various manifesto. A simple instance is the issuance of precious metal instruments and jewelry that have high intrinsic value. As an illustration, our observation of some precious metal retailers on the Tokopedia platform shows findings where gold bar products with the themes of "*Sekar Jagad*" and "*Wahyu Tumurun*" batik (Fig. 8) sold at a margin of 10% more than the market price of similar products, and there are still buyers. (see Table 5 for product price comparison). It can be imagined if the central and regional governments in Indonesia collaborate with Antam (or other credible companies) to issue limited edition precious metals for infrastructure funding purposes. Of course, from the retail sales margin, the government has the opportunity to fund several projects without disrupting regional/central spending.

Table 5. Gold Bar prices in Indonesia market (price per 31st January 2019)

Certification Issuer	Denomination	Retail Price	Retailer
Antam (IDX ANTM)	1 Gram	IDR 693,000	PT. Pegadaian
Antam (IDX ANTM)	2 Gram	IDR 1,353,000	PT. Pegadaian
Antam (IDX ANTM)	5 Gram	IDR 3,280,000	PT. Pegadaian
Antam (IDX ANTM)	10 Gram	IDR 6,478,000	PT. Pegadaian
Antam (IDX ANTM)	25 Gram	IDR 16,093,000	PT. Pegadaian
Untung Bersama Sejahtera (UBS)	0.5 Gram	IDR 366,000	PT. Pegadaian
Untung Bersama Sejahtera (UBS)	1 Gram	IDR 659,000	PT. Pegadaian
Untung Bersama Sejahtera (UBS)	2 Gram	IDR 1,302,000	PT. Pegadaian
Untung Bersama Sejahtera (UBS)	5 Gram	IDR 3,188,000	PT. Pegadaian
Untung Bersama Sejahtera (UBS)	10 Gram	IDR 6,345,000	PT. Pegadaian
Untung Bersama Sejahtera (UBS)	25 Gram	IDR 15,816,000	PT. Pegadaian
Antam (IDX ANTM)	1 Gram	IDR 699,000	Indo Gold
Antam (IDX ANTM)	25 Gram	IDR 16,039,000	Indo Gold
Antam (IDX ANTM)	50 Gram	IDR 32,200,851	Indo Gold
Antam (IDX ANTM)	100 Gram	IDR 64,323,100	Indo Gold
Antam (IDX ANTM) (Sekar Jagad edition)	10 Gram	IDR 7,250,000	Toko Agung Cikini
Antam (IDX ANTM) (Wahyu Tumurun edition)	10 Gram	IDR 7,250,000	Toko Agung Cikini
Antam (IDX ANTM) (Sekar Jagad edition)	20 Gram	IDR 13,600,000	Toko Agung Cikini
King Halim (Shuang Xi edition)	5 Gram	IDR 3,170,000	Toko Agung Cikini
King Halim (Dragon edition)	5 Gram	IDR 3,170,000	Toko Agung Cikini
King Halim (Balinese Dancer edition)	5 Gram	IDR 3,150,000	Toko Agung Cikini
King Halim (Masjidil Haram edition)	10 Gram	IDR 6,190,000	Toko Agung Cikini

From Table 5, it can be noted that the price of gold bullion in the Indonesian market is influenced by the issuer institution which certifying metal authenticity. This fact strengthens the results of our investigation that reputation vendors have a higher bargaining power towards the market. Besides, the availability of goods supply factor is another determinant ascertaining the high selling

prices of commodities in the market. In this context, Antam's precious metals with the theme "Sekar Jagad" and "Wahyu Tumurun" were made as a limited edition work of art. Limited production quotas will result in a scarcity of products on the market, which also increases consumer interest in having them [39].

The U.S Geological Survey report supports this fact [40], Where in 2016 and 2017 for gold commodities, Indonesia managed to produce as many as 80 tons regularly, this was only evaluated by one commodity, according to a report from the Indonesian Ministry of Trade [41], in 2014 Indonesia managed to export 8mm size Sea-shells on the global market with a value of US$ 28 million. Thus there are still many other opportunities that Indonesia can do for the needs of national development funding.

4.4. Investment engineering

For national development, all efforts must be made. Investment engineering is a way that must be taken to deal with the limitations of physical assets provided by nature or to deal with secondary financing which does not need to be shouldered by the government itself. It is time for the government to expand the concept of economic sharing with the community. Whether it is investment sharing or benefit sharing. The Government of Indonesia, through one of the State-Owned Enterprises, PT Waskita Karya has attempted to carry out economic sharing with the business sector by selling several toll road segments for the capital of other toll road project funding [42]. It is perfect and needs to be appreciated. However, the targeted market is still specific; in other words, only giant capital companies can invest in toll road projects; the administrative process will also be complicated. Therefore, involving all levels of society by eliminating administrative problems such as "*minimum investment of US$ 100*" or "*return on investment can be obtained two years later*" will undoubtedly be a powerful way to boost up the source of capital. Indonesian people need to learn from the video game industry to follow the way they do financial engineering as is done by steam-powered platforms from Valve video games company in the United States. They realize that the business model in the game industry that produces games and then sells them off the market will undoubtedly be very dull and make the market become saturated quickly. The way that Valve does to overcome this problem is by creating a community market.

The concept of a community market is to allow end-users to trade game items they obtain from gamification results (such as reward points when winning matches or leveling up) in steam-powered marketplace platforms. The concept of the community market applied by Valve, in our opinion has been successful in making playing video games not dull anymore. A win for all concept, both for platform owners, developers, and end-users. Furthermore, a member in the Reddit community forum once predicted the question related to "*how much money is produced by Valve only from every game item traded in the community market?*". Although the question looks like jokes, the money obtained by Valve is estimated to be very serious. Throughout 2012-2013. with a 15% discount on every successful sale of game items, it is estimated that from ± 58 million transactions and trading volume of ± US$ 39 Million, Valve managed to obtain ± US$ 6 Million [43].

5. Conclusion

The results of our investigation showed that there is good interest from the public regarding the decision to invest online, especially those vendors who have known their reputation. This situation is also strengthened by the fact that the products sold by reputable vendors have the opportunity to carry out a pricing premium strategy (selling at higher prices because vendors have a bonafide business image). The investment (financial) engineering method has also been very diverse, even in the video games industry. So we conclude that now is the exact moment for the Indonesian government to make creative breakthroughs to create national project funding opportunities. Because the available market for breakthroughs is not only from the local community but globally. If the government is willing to do that, then we believe that the Smart Cities concept will be successfully implemented in Indonesia and will bring prosperity to all people.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] R. Giffinger, "Smart cities Ranking of European medium-sized cities," October, vol. 16, no. October, pp. 13–18, 2007.

[2] I. D. G. of I. Application, Buku Panduan Penyusunan Masterplan Smart City 2017. Gerakan Menuju 100 Smart City. Jakarta, Indonesia, 2017.

[3] R. Mahesa, G. Yudoko, and Y. Anggoro, "Platform Ecosystems for Indonesia Smart Cities," in 2018 International Conference on Computer, Control, Informatics and its Applications: Recent Challenges in Machine Learning for Computing Applications, IC3INA 2018 - Proceeding, 2019, pp. 34–39.

[4] R. Mahesa, G. Yudoko, and Y. Anggoro, "Dataset on the sustainable smart city development in Indonesia," Data Br., vol. 25, pp. 1–16, 2019.

[5] A. Ramadhiani, "Pemerintah Habiskan Rp 17,27 Triliun demi Lahan Tol Trans Jawa," Properti Kompas.com, Jul-2017.

[6] E. Simorangkir, "9 Proyek Bendungan Jokowi Rp 3,85 T akan Rampung Tahun Ini," detikFinance, Jan-2018.

[7] M. Idris, "Berapa Biaya Bangun 2.428 Km Rel Baru di Kalimantan?," Detik Finance, May-2017.

[8] H. Kusuma, "Ini Daftar Lengkap 14 Proyek yang Dicoret Jokowi dari PSN," Apr-2018.

[9] T. Hamdani, "Saat Jokowi Coret 14 Proyek Strategis Nasional Senilai Rp 264 T," Dec-2018.

[10] M. Wareza, "SBR 004 Beri Bunga 8,05%, Kemenkeu Cuma Incar Dana Rp 1 T," CNBC Indonesia Investment, Aug-2018.

[11] H. K. Dewi, "Kuota Nasional SBR005 Tinggal Rp1,47 Triliun Lagi, Ayo Segera Pesan," Bareksa News Funds, 2019. .

[12] D. Cyr, "Website design, trust and culture: An eight country investigation," Electron. Commer. Res. Appl., vol. 12, no. 6, pp. 373–385, 2013.

[13] R. N. P. Atmojo, A. D. Cahyani, B. S. Abbas, B. Pardamean, Anindito, and I. D. Manulang, "Design of single user decision support system model based on fuzzy simple additive weighting algorithm to reduce consumer confusion problems in smartphone purchases," Appl. Math. Sci., vol. 8, no. 13–16, pp. 717–732, 2014.

[14] Y. Lie, R. N. P. Atmojo, H. H. Muljo, S. Purnamasari, and B. Pardamean, "A comparison of customer preference towards two different types of e-Tailing channel," in Proceedings of 2017 International Conference on Information Management and Technology, ICIMTech 2017, 2018, vol. 2018-Janua, pp. 171–176.

[15] R. N. P. Atmojo et al., "Rekening Bersama (RekBer) The Indonesian Escrow Service of Money," Int. J. Multimed. Ubiquitous Eng., vol. 11, no. 3, pp. 399–410, 2016.

[16] A. O. Ajayi, G. A. Aderounmu, H. A. Soriyan, and A. David, "An intelligent Quality of Service brokering model for e-commerce," Expert Syst. Appl., vol. 37, no. 1, pp. 816–823, 2010.

[17] V. U. Tjhin, R. Tavakoli, and R. N. P. Atmojo, "The Determinants Affecting E-Loyalty: Hospitality Industry in Indonesia," in 2016 2nd International Conference on Science in Information Technology (ICSITech), 2016, pp.

131–136.
[18] M. Turilli, A. Vaccaro, and M. Taddeo, “The Case of Online Trust,” Knowledge, Technology & Policy, pp. 1–13, 2010.
[19] D. G. Gregg and S. Walczak, “Dressing Your Online Auction Business For Success: An Experiment Comparing Two Ebay Businesses,” MIS Q., vol. 32, no. 3, pp. 653–670, 2008.
[20] R. N. P. Atmojo, Anindito, B. Pardamean, B. S. Abbas, A. D. Cahyani, and I. D. Manulang, “Fuzzy simple additive weighting based, decision support system application for alternative confusion reduction strategy in smartphone purchases,” Am. J. Appl. Sci., vol. 11, no. 4, pp. 666–680, 2014.
[21] C. W. Yoo, G. L. Sanders, and J. Moon, “Exploring the effect of e- WOM participation on e-Loyalty in e-commerce,” Decis. Support Syst., vol. 55, no. 3, pp. 669–678, 2013.
[22] E. Toufaily, N. Souiden, and R. Ladhari, “Consumer trust toward retail websites: Comparison between pure click and click-and-brick retailers,” J. Retail. Consum. Serv., vol. 20, no. 6, pp. 538–548, 2013.
[23] R. N. P. Atmojo et al., “Research Plan Development Concerning E-Image Impact Towards Online Purchase Intention And Premium Pricing Strategies In Indonesia Community Based Online Market,” in 2015 International Conference on Science in Information Technology (ICSITech), 2015, pp. 1–6.
[24] R. N. P. Atmojo and A. D. Cahyani, “Recommendations for Development of Online Bookstore in Indonesia: a Review From Writer’s Perspective,” J. Comput. Sci., vol. 10, no. 8, pp. 1522–1530, 2014.
[25] S. Vavilis, M. Petkovi?, and N. Zannone, “A reference model for reputation systems,” Decis. Support Syst., vol. 61, no. 1, pp. 147–154, 2014.
[26] L. Liu, X. Zhao, H. Wang, W. Song, and C. Du, “Research on Identification Method of Anonymous Fake Reviews in E-commerce,” TELKOMNIKA, vol. 14, no. 4, pp. 1510–1520, 2016.
[27] Y.-T. Chen and T.-Y. Chou, “Exploring the continuance intentions of consumers for B2C online shopping Perspectives of fairness and trust,” Online Inf. Rev., vol. 36, no. 1, pp. 104–125, 2014.
[28] D. Gefen and D. W. Straub, “Consumer trust in B2C e-Commerce and the importance of social presence: Experiments in e-Products and e-Services,” Omega, vol. 32, no. 6, pp. 407–424, 2004.
[29] A. Ogonowski, A. Montandon, E. Botha, and M. Reyneke, “Should new online stores invest in social presence elements? The effect of social presence on initial trust formation,” J. Retail. Consum. Serv., vol. 21, no. 4, pp. 482–491, 2014.
[30] D. Gefen, “Customer Loyalty in E-Commerce,” J. Assoc. Inf. Syst., vol. 3, no. 1, pp. 27–51, 2002.
[31] D. Gefen, “E-commerce: the role of familiarity and trust,” Omega, vol. 28, no. 6, pp. 725–737, 2000.
[32] Y. Hwang and K. C. Lee, “Investigating the moderating role of uncertainty avoidance cultural values on multidimensional online trust,” Inf. Manag., vol. 49, no. 3–4, pp. 171–176, 2012.
[33] H. Xu, R. Zhang, C. Lin, and W. Gan, “Construction of E-commerce Recommendation System Based on Semantic Annotation of Ontology and User Preference,” TELKOMNIKA Indones. J. Electr. Eng., vol. 12, no. 3, pp. 2028–2035, 2014.
[34] I. B. Hong and H. Cho, “The impact of consumer trust on attitudinal loyalty and purchase intentions in B2C e-marketplaces: Intermediary trust vs. seller trust,” Int. J. Inf. Manage., vol. 31, no. 5, pp. 469–479, 2011.
[35] E. Symeonaki, M. Papoutsidakis, D. Tseles, and M. Sigala, “Post-implementation evaluation of a University Management Information System (UMIS),” in Proceedings - 2016 3rd International Conference on Mathematics and Computers in Sciences and in Industry, MCSI 2016, 2017, pp. 14–19.
[36] W. W. Chin, “The Partial Least Square Approach to Structural Equation Modeling,” in Modern Methods for Business Research, 1998, pp. 295–336.
[37] K. K. Wong, “Partial Least Squares Structural Equation Modeling (PLS-SEM) Techniques Using SmartPLS,” 2013.
[38] W. Wandoko, B. S. Abbas, D. Budiastuti, and R. Kosala, “Online trust building through third party trust transfer and third party protection,” J. Phys. Conf. Ser., vol. 801, pp. 1–8, 2017.
[39] S. Worchel, J. Lee, and A. Adewole, “Effects of supply and demand on ratings of object value,” J. Pers. Soc. Psychol., vol. 32, no. 5, pp. 906–914, 1975.
[40] U. S. G. Survey, “Mineral Commodity Summaries 2018,” 2018.
[41] T. M. of T. R. of Indonesia, “Indonesian South Sea Pearls,” 2016.
[42] A. Basith, “Waskita Karya jual 18 ruas tol untuk menutupi utangnya,” Kontan Industri, 2018.
[43] Reddit, “Valve’s Cash Cow - The Steam Community Market.” 2013.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 222-225 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Stability of Ionized Air Using Density Functional Theory (DFT) for High Voltage Engineering

Fri Murdiya[1], Neni Frimayanti[2,*]

[1]*Department of Electrical Engineering Faculty of Engineering Universitas Riau Kampus Bina Widya, Jl. HR, Subantas KM 12.5 Simpang Baru Tampan Riau Indonesia*

[2] *Sekolah Tinggi Ilmu Farmasi (STIFAR) Riau, Jalan Kamboja, Pekanbaru, 28293, Indonesia*

ARTICLE INFO

Article history:
Received: 24 October, 2019
Accepted: 25 November, 2019
Online: 22 January, 2020

Keywords:
DFT
Ionization
Exitation energy
High voltage
Gas insulation

ABSTRACT

In high voltage engineering, air is the one insulation gas that can ionized cause by electrical discharge. Air consisted of many gases such as O2, N2, CO2, H2, etc. In order to study the characteristic of the insulation gas, we can apply density functional theory (DFT) to determine gas ionization level such as charge distribution and excitation energy. These data are shown the stability of gas molecule when ionization process occure cause by electrical discharge, ultraviolet light, etc. For application on ozone production by electrical discharge, the oxygen atom can transform to ozone need energy above the stabiity excitation energy of oxygen.

1. Introduction

In recent year, the high voltage engineering is focused on electrical discharge on the insulation dielectric in the form solid, liquid and gas. Base on our previous research, computational approach can be used to study the air failure in the high voltage which can be utilized to generate ozone from the ionization process of the oxygen atoms [1]. The ionization process has been checked using homo-lumo application. Some gases such as hydrogen, oxygen, nitrogen, carbon dioxide widely used as high voltage insulation. These gases are also used as an ionized gas for application in the high voltage like to produce plasma. This plasma was then used for physics, chemistry and medicine fileds. Recently, laboratory testing was required to determine the electrical strength (E) of that gas [2] or by kinetic modeling of the breakdown process on the basis of a microscopic electron-molecular cross section [3]. These methods are needed much more time and resource intensive and only can be applied to narrow gas.

There are some obvious advantages of the calculation method compared than traditional methods such as saving cost in term of time and effort. Moreover the calculation approach allows for identifying the atomistic processes of insulation gas, and also for exploring the intrinsic changes inside the sensing materials. Along with the development of computational materials science, the theoretical calculation methods can not only be used to provide the reasonable and profound explanations for the experimental results, but they are helpful for the design of new materials structures and thus new functionalities Previous reports reported that E is highly correlated with certain predictors which are simple functions of certain molecular properties [4]. These molecular properties are calculated using Density Functional Theory (DFT). Thus, E for molecules can be predicted by their own computational methods before testing their electrical and chemical behavior in experiments.

In this research, density functional theory (DFT) has been developed as reliable and effective approach for the computer simulation of molecular structure and also to measure the energy for chemical reaction. DFT is also can be used to search the efficiency and acuracy of the molecular properties, such as electronegativity, chemical potential and affinities [5]. The molecular properties of can be accurately described and calculated by manipulating the electron density and its fundamental quantity.

2. Research Methodology

2.1. Preparation of Molecular Gases

Preparation of molecular gases i.e. O2, N2¸ H2, CO and CO2 are beginned with skecthed of each molecular gases using ChemBioDraw ultra 13.0 for then it should save it in mdl. mol format. for then imported it into the Gaussian software.Finally, complete content and organizational editing before formatting.

*Corresponding Author Neni Frimayanti, Address : STIFAR Riau, Jalan Kamboja, Pekanbaru, Indonesia,, email: neninfrimayamnti@gmail.com

www.astesj.com
https://dx.doi.org/10.25046/aj050128

Please take note of the following items when proofreading spelling and grammar.

2.2. Density Functional Theory (DFT)

The initial structure of each gas O2, N2, H2, CO and CO2 are needed to obtain, density functional theory (DFT) is urgently needed to optimize the stability of these gases. DFT was runned using Gaussian 09 software with 6-311G/B3YLP was selected as basis set. Basis set is shape of atomic orbital that explained using wave function. TD-SCF was used as method to perform the Gaussian [6]. Stability of the calculation was explored the stability of the wavefunction computed for a molecular system. The stability of SCF solution for unknown systems should always be tested. The consideration for stability of unknown systems should be apply and may be tested in calculation using DFT methods as well.

3. Results and Discussion

Single particle are described about molecules or atom or also described how the chemical bonding between each atom to make a molecule. In this study, density functional theory (DFT) was applied to check the stability of gases and also to study shock induced process because they involve spatial and temporal scales that go beyond those attainable at the atomistic level. The stability of each gas such as N2¸O2, CO, CO2¸ and H2. base on the DFT calcution we will know charge distribution and excitation energy [7]. DFT calculation provides the energy of ionization potential for these gas elements.

Based on DFT simulation for these gases N2, O2, H2, CO and CO2, these gases are looks stable with the excititation energy of 0.56334 eV, 0.56334 eV, 0.49436 eV, 0.52185 eV, 0.53651 eV, respectively. Plate testing model (anode and cathode electrodes) was used for high voltage for then it can be simulated the excitation energy using the equation 1 (eq. 1) with d = 0.01 m. The results of the calculation of excitation energis is presented in Table 1.

$$W = F.d \quad (1)$$

Where, W is excitation energy (eV) and F is Coulomb force on electrical charges (N)

$$F = E \,.\, e \quad (2)$$

E is electric Field/ electrical strength (V/m) and e is electron

Initially, the DFT for each gas was evaluated using Gaussian 09 software package. Instead of DFT, it also can predict the spectrum UV-Vis for each gas. Based on spectrum UV-Vis, the information about energy excitation will be informed [8].

3.1. Nitrogen

The gas N_2 is seems to be stable at the wave length 89.17 nm, based on DFT calculation for the best pose selected with the following results; energy of excitation of 89.17 eV, energy for the ionization process of 108.98 au from orbital eight of -0.13492 to orbital seven of -0.42842. The ionization process of molecule N_2 is depicted in Figure 1.

Molecule consisted of atoms, atoms consisted of electrons, proton and neutron. Excitation of each electron in orbital atoms can cause the ionization process. The ionization process is stable if the range energy between two orbital less than 0.9 eV [9]. For gas N_2, there are two electrons was jumped from orbital eight (4p) into orbital seven (4s) with the range energy of 0.56334 eV. It is indicated that gas N_2 is stable.

Table 1. Gases excitation energy

No	Gases	W (eV)	E (V/m)
1	Molecule N_2	0.56334	56.334
2	Molecule O_2	0.56334	56.334
3	Molecule H_2	0.49436	49.436
4	Molecule CO	0.52185	52.185

There are some reasons to make the frontier orbital become so significant, they are: method for calculating the ionization energy and Koopman's theory. Based on Homo-Lumo calculation on our previous paper [10], it is provided the energy of ionization potential for gas elements. Furthermore, DFT calculation is needed to analyze much more about the orbital for basic gas.

Each molecule has orbital or molecular orbital (MO), it can be used to represent the regions in a molecule where an electron occupying that orbital is likely to be found. Molecular orbitals also can be used to predict the location of an electron in an atom. A molecular orbital can specify the electron configuration of a molecule: the spatial distribution and energy of one (or one pair of) electron(s). The molecule will stabilize if the electrons in the last orbitals have their paired up [11].

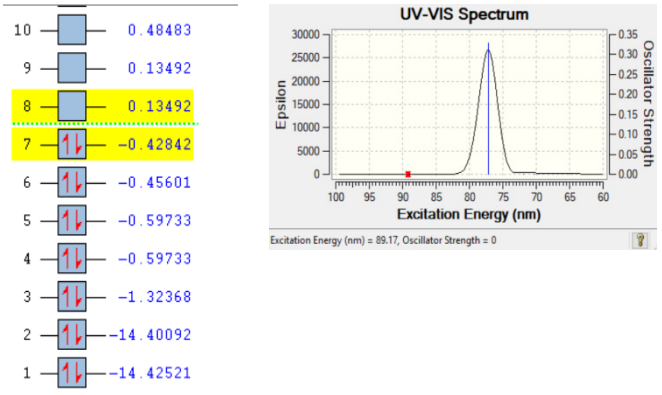


Figure 1: Ionization process of gas N_2

3.2. Oxygen

Generally, DFT can be defined as method for computational quantum mechanical modelling used in physic, chemistry or material science for investigating the properties of electron in a system, hence the name density functional theory comes from the use of functionals of the electron density [12].

Like gas N2, ionization of molecule O2 is looks stable since the excitation energy of -395143 eV with the wave length of

395143 nm, the best pose selected with the following results; energy for the ionization process of 108.98 au from orbital nine of 0.01523 to orbital eight of -0.05553. The ionization process of molecule O2 is presented in Figure 2. Stability of gas O2 is shown with there are two electrons was jumped from orbital nine (4d) into orbital eight (4p) with the range energy of 0.07053 eV.

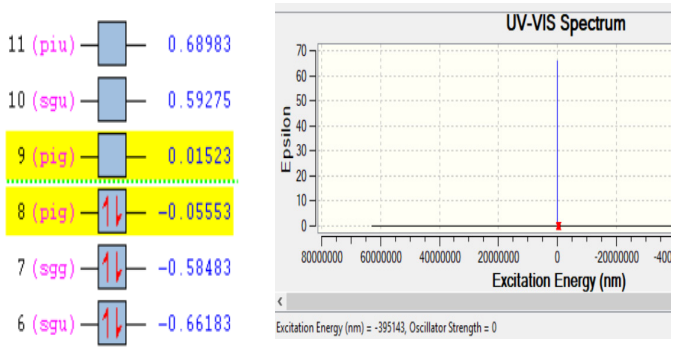


Figure 2: Ionization process of gas O_2

3.3. Hydrogen

The stability of molecule H_2 is also checked using the same way like molecule O_2 and N_2. Base on DFT calculation, it come out with the excitation energy of 88.8 eV at the wave length 88.8 nm. Two electrons were jumped from orbital two into orbital one with the energy of 0.49436 eV. Figure 3 is presented the ionization process of molecule H_2.

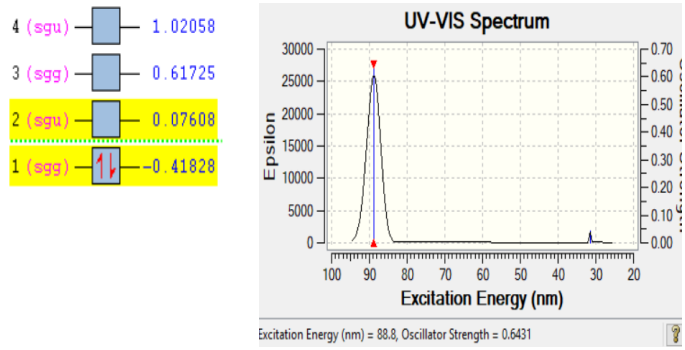


Figure 3: Ionization process of gas H_2

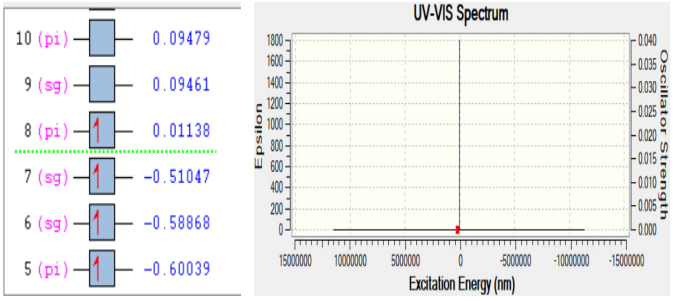


Figure 4: Ionization process of gas CO

3.4. Carbon Monoxide

Stability of molecule CO was checked using DFT. The first principles of DFT is to carried out the structural stability and also the electronic properties of a molecule. Base on DFT calculation, gas CO has the excitation energy of 167782 eV. Only one electron was excited from orbital eight into orbital seven with the energy of 0.52185 eV. The process of electron excitation for molecule CO is depicted in Figure 4.

3.5. Carbon Dioxide

DFT determined the stability of molecule CO2. Base on the DFT simulation, gas molecule CO2 is also looks stable with the excitation energy of 96.13 eV. There are two electrons were excited from orbital eleven into orbital ten with the energy of 0.53651 eV. Figure 5 is presented the ionization process of molecule CO2.

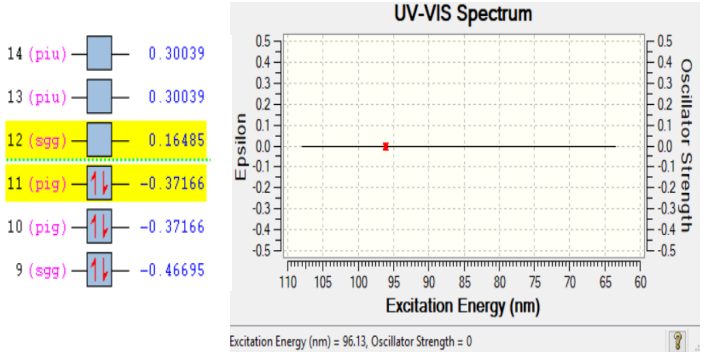


Figure 5: Ionization process of gas CO_2

4. Conclusion

There are several conclusions that can be concluded base on this research:

1. Density functional theory (DFT) has been developed as reliable and effective approach for the computer simulation of molecular structure.

2. The ionization process is stable if the range energy between two orbital less than 0.9 eV.

3. Gas molecule N_2, O_2, H_2, CO and CO_2 is stable with the excititation energy of 0.56334 eV, 0.56334 eV, 0.49436 eV, 0.52185 eV, 0.53651 eV

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We thank to Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) Universitas Riau for the financial support through Penelitian DIPA UNRI percepatan inovasi 2019, contract Number :1026/UN.19.5.1.3/PT.01.03/2019

References

[1] L. B. Affoue, B. A. Amon, D. Y.. Kicho, G. K. Mamadou, K. Soleymane, K. Mawa, R. N. Boka, S. B. El-Hadj, "Theoretical study by density functional theory method (DFT) of stability,tautomerism, reactivity and prediction of acidity of quinolein-4-one derivatives" Computational Chemistry. **6**, 57-70, 2018. https://doi.org/10.4236/cc.2018.63005

[2] G. Biasutti,'Neue Isoliergase für gekapselte Hochspannungsanlagen". Ph.D. Thesis, Diss. ETH Nr. 7683, 1985

[3] R. E. Robson, P. B. Nicoletopoulos, R. D. Li, "Kinetic theoretical and fluid modelling of plasmas and swarms: the big picture" Plasma Sources Science and Technology, **17**, 024020, 2008. https://doi:10.1088/0963-0252/17/2/024020

[4] M. Rabie, D. A Dahl, S. M. Donald, M. C. Reiher, M. Franck,. "Predictors for Gases of High Electric Strength". IEEE Transactions on Dielectrics and Electrical Insulation, **20** (3), 856–863, 2013. https://doi: 10.1109/TDEI.2013.6518955

[5] M. S. Jorge, "An introduction to density functional theory in chemistry" 2018. Theoretical and computational chemistry, **2**, 1-27, 2018. https://doi.org/10.1016/S1380-7323(05)80031-7

[6] M. J. Frisch, G. W. Trucks, H. B. Schlegel, "Gaussian 03, Revision A.1" Gaussian Inc., Pittsburgh, 2003

[7] A. Nathan, M. Guy, "Density Functional Theory — an introduction" American Journal of Physics. **68** (1), 69-79, 2000. https://doi.org/10.1119/1.19375

[8] K. Erick, L.Vladimir, "Study of gas-phase O–H bond dissociation enthalpies and ionization potentials of substituted phenols – Applicability of ab initio and DFT/B3LYP methods" Chemical Physics, **330**, 515–525, 2006. https://doi.org/10.1016/j.chemphys.2006.09.026

[9] J. Ahmad, J. Li, M. Hussain, M. T. Riaz, M. A. Akbar, and M. Z. Khan, "Study of Physio-chemical Properties of POSS/Mineral Oil based Nanofluids," in 2018 International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET), 2018, pp. 1–5. https://doi.org/10.1109/PGSRET.2018.8685972

[10] M. Kurt, T. R. Sertbakan, M. Ozduran, "An Experimental and Theoretical Study of Molecular Structure and Vibrational Spectra of 3- and 4-Pyridineboronic Acid Molecules by Density Functional Theory Calculations". Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, **70,** 664-673, 2008. https://doi.org/10.1016/j.saa.2007.08.019

[11] Q. Hai, L. Wenhao, W. Xiaoxing, C. Wei, D. Jun, H_2S and SO_2 adsorption on Pt-MoS_2 adsorbent for partial discharge elimination: A DFT study. Results in Physics, **12**, 107-112. 2019. https://doi.org/10.1016/j.rinp.2018.11.035

[12] Z. Yubo, W. James, X. F. Bing, S. Jianwei, "Subtlety of TiO_2 phase stability: Reliability of the density functional theory predictions and persistence of the self-interaction error" The Journal of Chemical Physics. **150** (01), 014105. 2019. https://doi.org/10.1063/1.5055623

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 226-232 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Framework for Automation of Cloud-Application Testing using Selenium (FACTS)

Md Nurul Islam*, Syed Mohammad Khurshaid Quadri

Department of Computer Science, Jamia Millia Islamia (A Central University), 110025, India

ARTICLE INFO

Article history:
Received: 10 December, 2019
Accepted: 08 January, 2020
Online: 22 January, 2020

Keywords:
Test
Automation
Framework
Selenium

ABSTRACT

A framework is an amalgamation of integrated tools, libraries, utilities and its coordination to interact with the other automation components. The motive of designing a test automation framework is to implement uniform standards towards automation throughout the organization to achieve the desired outcomes. Test automation framework is required to maintain the standardization of the activities performed to manipulate the operations (addition, modification and deletion) on the test scripts and functions easily, to provide the scalability and reliability. Tools, process and testing team are the three key player of test automation. The proposed framework is based on the consideration of test environment (includes language binding, IDE, automation tools and testing framework). Selenium WebDriver is used as web application automation framework to execute test across multiple browsers that supports multiple operating system with variant programming language. Selection of WebDriver automation tool for this framework is due to their internal architecture to directly communicate with browser for fast execution. The combination of test automation tools (i.e. Java, Eclipse, Selenium and TestNG) makes JEST for designing test automation framework. Both Positive and negative test must be performed to verify functionalities of applications to handle unusual exceptions.

1. Introduction

Software testing is an evaluating as well as mitigating process, as it evaluates the functionalities of application to meet out the specific requirements of software product, at the same time it also identify the defects in product to mitigate the risk of software failure by find bugs and errors in the program.

"Framework for Automation of Cloud-application Testing using Selenium (FACTS)" is almost self-explanatory, though this paper may also explore the terminology explicitly. This paper may discuss about, why do we require test automation framework, how the software testing framework may organize the test automation, how do we select best appropriate software testing framework, what are the challenges being faced by the software automation team to design and deploy the test automation framework, what are the potential benefits of using software testing framework.

Why selenium for automation? Selenium is powerful open source tool for test automation framework independent of language binding & development platform that support functional testing across the browsers.

A framework is a combination of various forms of principles, guidelines, processes, concepts and tools etc. followed by testers to take advantage of it, to automate the application. Test automation framework decides the success or failure of any software automation project. Test automation is testing process of automating test with the help of automation tools to find out defects in the application. Test automation reduces test execution time and testing cost along with the less human interaction with the system. Test script has been used to execute test multiple times, when required. Test script must consider, what to automate or what not to be, like functionalities that are going to be change in near future or not, whether script is applicable for regression testing or not, whether test cases takes more time to execute/execute with errors or operated over huge amount of input data etc.

2. Test Automation Framework – An Understanding

Test automation framework is set of guidelines/rules followed while writing test script to generate test cases as per the

* Md. Nurul Islam, Department of Computer Science, Jamia Millia Islamia, New Delhi, India 110025, 9818933615 & mnislam@jmi.ac.in

requirement of test scenarios. A framework is an amalgamation of integrated tools, libraries, utilities and its coordination to interact with the other automation components to execute the test scripts and to capture the outcomes in systematic and beneficial manner to test a specific product. If a testing activity automated without having a test automation framework, it is obvious to raise the question that "how reliable the method would be? And how quickly they can produce the same output?"

The motive of designing a test automation framework is to implement uniform standards/policies towards automation throughout the organization to achieve the desired result. A test automation framework can be considered effective if the result of test automation is easily understandable to end user, otherwise executing high volume of test cases are of little or no use for end user. The automated test generated report must include the point where the application gets failure along with test data being used for automation.

2.1. Why do we require test automation framework?

Test automation framework is required to maintain the standardization of the activities performed to manipulate the operations (addition, modification and deletion) on the test scripts and functions easily, with the aim to provide the scalability and reliability. If a tester automate testing activities without considering a framework, it is obvious to ask how reliable the method would be. Can the tester re-generate the same result? An automation framework sort out all such problems by providing a standard guideline to leverage reusable components to attain the desired & reliable results rapidly. There are various factors for which an automation framework required [1].

- The Ease of use (Reusability of test script) to maintain the test script.
- To maintain reliability and consistency follow specific guidelines.
- To maintain scalability with changing requirement.
- To improve the efficiency of testing activities
- To provide easily understandable reports of automated results.

2.2. How do we select best appropriate test automation tool?

Selecting best appropriate tool for automation is one of the basic tasks for organization. The impact of selecting best automation tool may greatly affect the automation process which may provide sophistication in automation. While selecting tools for automation, certain conditions must keep in mind which makes sense for automation, like how frequently the repetitive test may occur, how often the regression testing is required etc. There are certain criteria to evaluate and select the automation tools [2] as mentioned below.

- Whether the tools fulfill the testing requirements. Is it compatible with project environment & technology that support objects in application?
- Which testing level and types does the tools support like (system, functional, performance) testing. Select tools which support maximum testing type with the ability to automate complex requirements.
- Whether the tools support all major browsers/OS or portable devices in which the application runs?
- Whether the estimated budget for purchasing automation tool affordable for organization?
- Whether the organization has skilled resources that can understand the language the tools support?
- Does the tool have powerful mechanism for graphical reporting?

While selecting a tool comprehensive understanding of its integration, reporting and ease of use with other automation tool must be executed. The standard automation testing framework should be application & tool independent & loosely coupled, it means that if the organization decides to migrate from one automation tool to other, it need not required to create a new test automation framework from scratch.

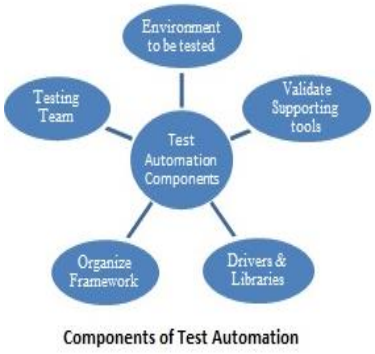


Figure 1: Components of Test Automation Framework

2.3. Components of test automation framework [3]

1) Environment to be tested: While designing a test automation framework all the test related tools, equipment, scripts and procedures must be verified and validated to ensure the system is stable to perform the desired task. This would be the first step towards the environment setup for the test automation. Until the test environment is not supposed to work accordingly the test automation may not be cost effective. Thus all the coordinating tools that support the environment must be stable and work together smoothly as a whole prior to start the test automation.

2) Validate supporting tools: While writing test scripts various supporting tools are required to organize the test. All such type of tools are properly investigated and independently monitored for expected contributory result as a unit, like test automation tool, bandwidth monitoring tools used for traffic management, other supporting tools such as requirement analysis, configuration management tools, test factory, defect tracking. A traffic monitoring and controlling activity of network is important to improve performance, efficiency and security of network resources. Network accommodates different kind of data traffic, identification of such traffic or congestion may lead the network administrator to optimize the network. Various types of network traffics are [4].

Table 1: Different Forms of Network Traffic

Traffic Type	Problem	Solution	Example
Busty Traffic	Consume high bandwidth starve applications	Fix constraint to limited bandwidth access	Download of FTP, Video Content
Latency Sensitive Traffic	Susceptible to compete for bandwidth results in poor response time	Fix min. & max. bandwidth range based on priority	Streaming application, VOIP, Video Conferencing
Interactive Traffic	Susceptible to compete for bandwidth results in poor response time	Prioritize over less essential traffic	SSL transactions, IM, Telnet Session
Non-Real Time Traffic	Consumes bandwidth during business hours	Schedule bandwidth during non-business hours	Email, batch processing applications

Configuration management tools are used to implement the changes (add, remove and update) occur during the test automation framework design to all associated system by implementing and enforcing to adopt such possible change. Test factory is an extensive use of test automation that minimizes the cost of testing and error rate along with supporting rapid scalability. Defect tracking is tracking the logged defects from beginning to end by testing, inspection or from customer feedback in a product with the aim to provide latest version of the product to fix the defect.

3) Drivers & Library: Different browsers require their own driver to communicate and control with the WebDriver. Some of the browsers are OS dependent but Firefox and Chrome are easily available across the platform. Download and unpack the drivers and add the location to you system PATH variable. The browser driver interacts with the respective browser by establishing a secure connection. While a test script is executed with WebDriver environment various activities performed in the background like:

- For every selenium command an HTTP request is generated to communicate with the browser driver.
- Drivers receive the HTTP request trough HTTP Server.
- HTTP Server decides the execution of instruction on the browser.
- As HTTP Server receives the execution status, it automates the script.

Browser automation library has been used to interact with other task required for automation. Library works as basic building block to establish a successful test environment to execute the test script by providing the required jar files, libraries, drivers and other supportive files.

Steps to add the Selenium Library to the project:

- Download & extract current release of Java Selenium library.
- Open the Eclipse IDE, go to the "Project→Properties"
- New dialog box open, Click on " Java Build Path" in the left pan, the to the register "Libraries"
- Click on "Add External JARs" in the right pan.
- Move to selenium library downloaded folder to select all .jar files by clicking "Open" button.
- Repeat this for all .jar files in the sub folder libs as well.

Once the library is configured with the project it can be reusable with the different tests, such utilities can perform different tasks for different test cases.

4) Organize Framework: Selenium is automated testing framework for web applications. To organize the framework following major steps can be taken into consideration [5].

- Select Programming Language: Select appropriate programming language as per the need of application under test and the developer/tester going to use the framework to write tests?
- Select Unit Test Framework: To design a framework a unit test framework is required. TestNG Compatibility with JAVA is highly recommended as it eliminates the limitation of earlier framework with the ability to write more flexible and more powerful tests. Other features TestNG are: annotation, grouping, sequencing and parameterization etc.
- Design the Framework Architecture: The architecture of framework is designed in such a way that it can be easily maintainable, scalable and sustainable in nature.
- Build the SeleniumCore component: SeleniumCore is designed to communicate between the web elements and browsers instances, it is used to create and destroy WebDriver objects, since it is not being taken into consideration by test writers.
- Build SeleniumTest Component: SeleniumTest components are all those test cases going use the classes provided by SeleniumCore. The design pattern used is Page Object Model (POM). It means for every page in the web application their corresponding a Page Object for them to organize the UI element in the pages.
- Select a Reporting Mechanism: Unless we get useful insights from the test result to take meaningful corrective action, test automation is of no use. A good report must include details about number of pass or failed test cases, passing rate, execution time and reason why test cases failed.

5) Testing Team: Successful test automation depends on efficient testing tools, standard testing process and skilled testing team to execute assigned responsibilities. Tools, process and testing team are the three key player of test automation. Software testing team is mainly engaged for test automation to utilize dedicated automation testing resources. The test team ensures that

the success of testing and automation effort based on the roles, duties and skill they provide.

- ✓ Manual Tester – Design and develop manual scenarios from requirement analysis.
- ✓ Automation Tester – Design and develop automation script.
- ✓ Lead Automator- Design and develop project automation architecture. Provide training and guidelines to testing team and collaborate to achieve the testing goal.
- ✓ Automation Environment Expert- Design, configure and update test environment and automation software.

Apart from the above role and responsibility testing team manages test platform, test tools, test cases and also provides tutorial to the users.

2.4. Characteristics of test automation framework

- If test automation framework is independent of Operating System, Programming language, tools and browsers, it would be considered as standard framework.
- The framework must be easily understandable without having programming skills.
- The framework must be easily maintainable and extendable.
- The Framework must execute test cases automatically, which includes test scenarios, test data, report generation, error handling and reporting final result.
- The framework must be able to execute tests without human intervention.
- Every test executed, must assign status of either Pass or Fail, failed test cases must include the failure data, the point of failure and other short description for failure.
- The framework should have meaningful logging and reporting structure.

2.5. Types of automation framework [6]

1) Linear Scripting Framework - This framework is also known as 'Record and Playback' framework as it generates the test scripts based on simulated record and playback method. It is not useful to run with multiple data set i.e. lack of reusability. This framework is suitable for small size applications. It works for individual test cases because it records the test script for particular scenarios.

2) Modular Testing Framework - In this framework, the complete application is divided into number of smaller modules and test script is generated individually. Finally the results of individual modules are compiled together by the master script to achieve the required target.

3) Keyword Driven Testing Framework – This framework is based on the keywords or actions for executing functions or methods. The test script is totally keyword/action dependent stored in tabular format in excel sheet, so it is also known as table-driven testing. The main class maintains the keywords and their corresponding action in the external table.

4) Data Driven Framework – Data Driven Test Automation framework is based on bifurcating the test data from test script. Test data is stored in separate files/format, test scripts connect and execute with different sets of test data. Due to having variant combination of test data set in separate table, this framework covers more test cases with flexibility in execution of test by changing input test data. This framework requires the programming skill to develop the test script.

5) Hybrid Driven Testing Framework – This framework is combination of two or more framework to leverage the strengths and benefits of each of the framework individually and to produce a better test environment overall.

6) Behavior Driven Development Testing Framework – The Purpose of using this framework is to share the process and tools and to communicate amongst stakeholders, with the aim to deliver quality product. BDD test cases are being written in simple language that can be easily understandable, which contains the details of how application behaves? BDD provides the platform for both technical and non-technical team to work together. A test case written in BDD follows the format 'Given-When-Then'. Given a certain scenarios, When an action takes place, Then this should be the outcome.

3. Related Works

Selenium WebDriver is powerful tool for test automation that directly interacts with web browsers, it follows the page object pattern to access the web page to reuse the test script to write the test cases. The page object pattern is used to reduce the coupling between web pages and test cases, which improves test suite maintainability [7]. UI locators are used to locate the web elements (Text box, button, checkbox, etc.) on the web page. Different identifiers are used to locate the web elements like (ID, link, XPath, link Test etc.), ID Locators have better result than XPath Locator, Combination of ID locator or XPath Locator with Link Text have improved the result but still not conclusive [7].

WebDriver tool can be used to write test cases easily and efficiently, it can helpful to develop and analyze the test cases using screenshot property, at the same time it is easily maintainable, reusable and extendable with the new requirement. Since the selenium WebDriver does not support any built in mechanism to generate test report, for that purpose TestNG is used as a testing framework. TestNG is compatible to perform unit, functional and integration testing. The author has generated a customized test report to analyze the failure test cases using screenshot [8].

WebDriver is cross platform testing framework that provides dynamic characteristics of web page, it means page reload is not required, if page elements gets changed. Selenium WebDriver execute faster as it directly interacts, run and control browsers using own control engine [9]. Selenium WebDriver is not equipped with reporting mechanism, TestNG generate report on behalf of selenium result. Such report contains execution time, method and detailed status of test script of running test suite "Passed or Failed" [10].

Table 2: Comparisons of Automation tools

Products & Parameters	Selenium	Katalon Studio	Test Complete	QTP/ UFT	Soap UI
Ease of Configuration & use	Advance skill required for configure & execution	Easy to configure & execute	Easy to configure. Training required for tool operation	Expertise require for configuration. Training required for tool operation	Easy to configure & execute
Licences required	Open Source & free to use, Free of Charge	No licence required. Not open source but free	Commercial tool, Licence required	Commercial, It is HP licenced product avail as single/Concurrent user	SoapUI is an open source. SoapUI Pro is the commercial.
Service support	Expert or group professional support	User friendly interface	Community Volunteer support	HP professional support	SoapUI pro Customer get support from forum or wen form
Operating System Support	Ms Windows, Apple OS X, Linux, Solaris	Ms Windows7/8/10, Apple OS X, Linux	MS Windows10/8.1/8/ 7/ vista, Server16/12/8	Only Windows Windows7/8/8.1/ XP/Vista/ 2003/2008	Ms Windows, Apple OS X, Linux
Scripting Language	Java, C#, Perl, PHP, Python, R, Ruby, JavaScript	Groovy, Java	JavaScript, C#, Python, VBScript, JScript	VB Script, Java Script, Window Shell Script	Groovy
Browser Support	Firefox, IE, Chrome, Safari, Opera, Phantoms, HtmlUnit, Android, iOS	IE9/10/11,Firefox, Chrome, Opera, Safari, MS Edge, Remote, Headless Browser	MS Edge,IE10/11, Chrome, Firefox	IE(V6-V11), Firefox(V3-V31),Chrome(V12-35),Safari6/6.1/7	SoapUI Default Browser
Device/Technology Support	Web apps	Web/API/Mobile/ Desktop apps	Desktop, Web & Mobile apps	Web/API/Mobile/ Desktop apps	API/Web apps
Testing Framework	RSpec (Ruby), Test::Unit (Ruby), Unittest (Python 2), JUnit 4 (Java), TestNG (Java), NUnit (C#)	Katalium Framework with TestNG	Application's UI	Build-in Feature for framework, like Key-word driven, Data Driven, Hybrid	Application's GUI
Type of test Performed	Functional Testing	Cross Browser Testing	Unit Testing, Functional Testing	Functional, Regression & Service Testing	Functional. Load & compliance Testing
Initially launched	2004	2015	1999	1998	2005

Comparisons of automation tools [11]

4. Framework Design for Automation of Cloud-Application Testing Using Selenium

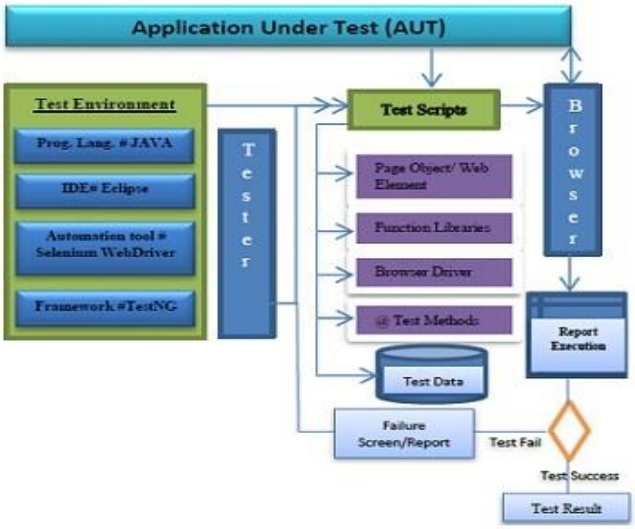


Figure 2: Architecture of Proposed Framework (FACTS)

5. Proposed Work

The framework (FACTS), we have designed is based on the consideration of test environment (includes language binding, IDE, automation tools and testing framework), types of Application under Test (AUT) with the compatibility of web browsers. The architecture of the proposed framework (Figure 2) includes following tools, functions, method and libraries etc.

- Programming Language # Java
- IDE #Eclipse
- Automation Tool # Selenium WebDriver
- Framework # TestNG
- Test Script
- Page Object /Web Element
- Function Libraries
- Browser Driver
- @Test Methods
- Test Data

The proposed framework may utilize the following tools, functions, method and libraries for designing automation framework:

- *Programming Language #Java*: Why Java is selected for this test automation framework? Java is a programming language which supports multiple operating systems. It is platform independent language, support Object Oriented features. Other important characteristics are simple, robust, Secure, Multithreading, distributed, portable to use and free of cost availability, open source and licence free.
- *IDE #Eclipse*: Eclipse is an Integrated Development Environment (IDE) tool most widely used to develop applications in Java and other programming languages

like C, C++, Perl, PHP, Python, Groovy, Ruby, R etc. Eclipse development environment is customizable as per the workspace and plug-in required. It is free to download and install.

- *Automation Tool #Selenium WebDriver*: Selenium is widely accepted open source test automation tool to test web-applications. Selenium WebDriver API is easy to explore and understand which support tester to easily read and maintain the test. Selenium WebDriver is web application automation framework that execute test across multiple browser, supports multiple operating system, support different programming language also. WebDriver executes faster due to their internal architecture to directly communicate with browser. When we execute a selenium command, an HTTP request created to communicate with particular browser driver, the browser driver then uses HTTP services to get HTTP requests, to establish the selenium command. Since WebDriver does not provide any build in facility to automatically generate Test result, we have incorporated TestNG as test reporting tool.
- *Framework #TestNG*: TestNG evolved as a testing framework to overcome the shortcomings of Junit and NUnit. It is an open source used for test automation framework. Where NG stand for Next Generation. According to some survey 76% tester uses Java programming along with TestNG framework. TestNG also provides an option to execute failed test cases only, it creates a file named testing-failed.xml for failed test cases in output folder. If the tester is interested to see only failed test cases they can execute this .xml file. There are various features due to which TestNG is in demand:
    - TestNG covers a broad range of test at different levels started from unit testing to system testing.
    - It supports annotations to create test cases easily.
    - It introduces 'test groups' to group test cases with prioritization, if required.
    - It supports parameterization and parallel test execution, it also execute multiple programs/classes using .xml.

The combination of test automation tools used for proposed framework has gained immense popularity among test automation tools as language, IDE and framework. The combination of the tools (i.e. **J**ava, **E**clipse, **S**elenium and **T**estNG) makes JEST for design and develop test automation framework.

- *Test Script*: Test script is set of instructions execute to test the system functions on the application under test for expected results. Test script mainly includes detailed descriptions of actions with required data to perform test. Test script follows the sequence to explore how to use the program and the order in which it is being utilized/executed, by pressing which button to carry out particular action in the program along with the stepwise expected results to observer the change in UI. Test script has a drawback with project which changes their requirement frequently, so the testers have to continuously update the script to fulfill the requirements. Since the test script is designed to test a particular task frequently, if bugs occur in the code the result would be same every time until the tester change the conditions.
  Page Object /WebElement: A Page object is an object oriented class through which AUT interface the web page. It plays a key role to identify the element on the web page, at the same time it reduces the code duplication by enhancing the mechanism of test maintenance. Whenever a test needs to communicate with UI of the page they use only methods of page object class. In case the UI of the page gets changed, only test script within the page object needs to be change. All possible user interactions can be used as a method on the class, ex. *clickLoginButton ();*
  setCredentials (user_name, Password);
  Selenium encapsulates elements on the page as an object of WebElement. WebDriver uses various techniques to identify the element on the webpage based on the its properties like ID, Name, Class, XPath, TagName, CSS Selectors, Link Text etc. WebDriver uses method like findElement() to find single web element and returns WebElement object. The method findElement () uses various locators like By.id (), By.name (), By.XPath () etc. to locate the element on the page by its properties.
  Ex. WebElementemail_id=driver.findElement (By.id ("User_Name")); // to get the web element for email address text field.
  Selenium WebDriver uses *findElement(By.Locator())* method to find the element on the page. Here findElement method uses locator/object as 'By' identifier.
- *Function Libraries*: One of the major purposes of designing a framework is its reusability, if any functionality is used frequently, such function can be written at one place and called multiple times for other test case. It is known as generic function. User defined functions are created based on their own requirement used with help of calling the appropriate package file. Function libraries are used to maintain the logic which is being used to write automation test scripts and called with their name. Methods are also known as functions.
- *Browser Driver*: Selenium WebDriver is web Automation tools that help to run the test across multiple browsers like FireFox, IE or Chrome. Corresponding browser driver is required to run the test in particular browser. Selenium launches corresponding browser as

per the script to execute test steps. To initialize WebDriver we can write code as :

WebDriver driver = new FireFoxDriver(); // where WebDriver works as interface and driver is reference variable for interface. Object can be associated with the instance of class FireFoxDriver//

When we execute a selenium command (driver.getTitle();) an HTTP request will be created to communicate with particular browser driver, the browser driver then uses HTTP service to get HTTP requests to establish the selenium command. All the built-in methods of corresponding drivers can only be accessible after configuring the driver.

- *@Test Methods*: Why do we use @ annotation before writing test method. Annotations in selenium are used to bind the associated method going to execute. Test annotation @ always used before every method, if this test annotation '@' is not prefixed before the test method, that method would not be considered as part of test code to be executed. @Test indicates that the annotated method is a part of test case serves as the main method [12]. Some other annotations used with TestNG are:

  @BeforeSuite: This method will execute once before all tests in this suite have run.

  @BeforeTest: This method will execute before any test method belonging to the classes inside the <test> tag is run.

  @BeforeClass: This method will execute only once before the first test method in the current class is invoked.

  @BeforeMethod: This method will execute before each test method.

  @AfterMethod: This method will execute after each test method.

  @AfterClass: This method will execute only once after all the test methods in the current class have executed.

  @AfterTest: This method will execute after all the test methods belonging to the classes inside the <test> tag have executed.

  @AfterSuite: This method will execute only once after all tests in this suite have executed.

- *Test Data*: Test data is actual input used to test the application. Test data would be used for positive or negative testing. Positive data is used to verify the functionality of application to meet the expected results. Negative data is used to test the ability of program to handle unusual, exceptional or unexpected input with its corresponding outcome [13].

6. Conclusion

In-depth planning, process and efforts are expected while design and develop a test automation framework. Selecting right automation tools plays a pivotal role to accomplish the automation task. Selenium WebDriver has been used to test web application end-to-end. WebDriver is flexible to integrate with variant development environment, tools (test, reporting and automation), OS, browsers and methods to design a test automation framework. Various types of test automation frameworks are available to design and develop test automation. The proposed framework (FACTS) is designed to work with Data Driven Framework as well as Hybrid Driven Testing Framework. A Standard automation tool is expected to execute its integration with other automation tools and reporting mechanism. Positive and negative test must be performed to verify the functionalities of application and to check the ability of program to handle unusual exceptions respectively.

References

[1] "A Guide To Automation Frameworks |Smartsheet" . Smartsheet, 2019, https://www.Swmartsheet.com/test-automation -frameworks-software.

[2] K. Kunal, and Mohammad Saad. "Test Automation Tool Selection Criteria And Comparison Matrix (A Complete Guide)". Softwaretestinghelp.Com, 2019, https://www. Softwaretestinghelp.com/automation-testing-tutorial-4/.

[3] Genius, Software. "Six Major Components of A Test Automation Framework - Software Testing Genius". Software Testing Genius, 2019, https://www.softwaretestinggenius. com/six-major-components-of-a-test-automation-framework/.

[4] "Network Traffic Management | Ipv6.Com". Ipv6.Com, 2019, https://www.ipv6.com/applications/network-traffic-management/.

[5] Logigear.Com, 2019, https://www.logigear.com/blog/test-automation/ building-a-selenium-framework-from-a-to-z/.

[6] "Types Of Test Automation Frameworks | Software Testing Material". Software Testing Material, 2019, https://www. softwaretestingmaterial.com/types-test-automation-frameworks/.

[7] Leotta, Maurizio et al. "Comparing The Maintainability Of Selenium Webdriver Test Suites Employing Different Locators: A Case Study". ACM, 2013, doi:http://dx.doi.org/10.1145/2489280.2489284.

[8] Gojare, Satish et al. "Analysis And Design Of Selenium Webdriver Automation Testing Framework". Procedia Computer Science, Vol 50, 2015, pp. 341-346. Elsevier BV, doi:10.1016/j.procs.2015.04.038.

[9] K, Revathi, and Janani V. "Selenium Test Automation Framework in On-Line Based Application -IJARSE". International Journal of Advance Research In Science And Engineering, Vol. No.4, no. Special Issue (02) ISSN-2319-8354(E), 2015, pp. 55-63.

[10] Singla, Sherry, and Harpreet Kaur. "Selenium Keyword Driven Automation Testing Framework ". International Journal Of Advanced Research In Computer Science And Software Engineering, Volume 4, Issue 6, 2014, pp. 125-129.

[11] Nurul, Md., and S.M.K. Quadri. "Software Testing Approach for Cloud Applications (STACA) – Methodology, Techniques & Tools." 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Jan. 2019, 10.1109/confluence.2019.8776915.

[12] "Selenium WebDriver Tutorial | TestNG For Test Case Management And Report Generation | Edureka". Edureka, 2019, https://www.edureka.co/ blog/selenium-webdriver-tutorial.

[13] "Test Data Generation: What Is, How To, Example, Tools". Guru99.Com,2019,https://www.guru99.com/ software -testing-test-data.htm

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 1, 233-248 (2020)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

A Survey of Smart Hydroponic Systems

Falmata Modu[1],Adam Adam[*,1], Farouq Aliyu[2], Audu Mabu[1], Mahdi Musa[1]

[1]*Department of Computer Science, Yobe State University, 620242, Nigeria*

[2]*Department of Computer Engineering, King Fahd University of Petroleum and Minerals, 34464, Saudi Arabia*

ARTICLE INFO

Article history:
Received: 03 December, 2019
Accepted: 04 January, 2020
Online: 29 January, 2020

Keywords:
Hydroponics
Internet of Things
Smart Cities

ABSTRACT

The explosion in human population has left researchers scrambling for solutions on how to feed the world. Furthermore, rural-urban immigration has on the one hand left the farms in the rural areas devoid of farmers and on the other hand has left the urban areas over-populated. Hydroponics is a form of agriculture where crops are grown without soil. This technique allows the farms to follow the farmers to the urban area. In addition, the fact that no soil is needed, allows hydroponic system to be stacked vertically (also known as vertical farming) to save space. The final frontier in hydroponics is automation. It will allow one farmer to work more than one job and cultivate more than one farm simultaneously. This paper provides a comprehensive survey on smart hydroponic system developed to date.

1 Introduction

The soil is a valuable component of agriculture [1]: it provides support for the plants, it also provide nutrient to the plants and the soil provide a home to some of the microbial organism that forms a symbiosis relationship with the plants. However, all these ingredients can be provided with hydroponics. Hydroponics is the process of growing plants without soil [2, 1]. Evidence of hydroponics was found in the Egyptian wall painting [3]. There are many benefits to hydroponics [1]: 1) it does not require soil, 2) it is faster than traditional farming, 3)it requires less space and can be grown in any location, 4) it is unaffected by seasonal change, 5) little or no pesticides and herbicides are needed 6) Plants get complete range of nutrients they need at the quantity they need it, 7) Plants are protected against diseases and pests, 8) It can be used to isolate crops during experiments [4, 5, 6, 7].

Figure 1 shows the components of a hydroponic system: First there is a need for a location or growing area where the system is going to be installed since hydroponics require only water any space could be used for it. The reservoir is a container that stores the nutrient solution used by the system. Nutrients in a good hydroponic system must contain the optimum level of; oxygenation, salinity, pH, and conductivity of nutrient solution [8]. The hydroponics fertilizers contain six essential nutrients: N, P, S, K, Ca and Mg, which are fed to the plants in form of mutual ratio of anions: $NO_3^-, H_2PO_4^-$ and SO_4^{2-}, and the mutual ratio of cations K^+, Ca^{2+} and Mg^{2+} [9]. Several growing medium like a wood chip can be used with water to form a hydroponics system [10]. Light is necessary for photosynthesis, in close/indoor hydroponic systems light-emitting diodes (LEDs) and other sources of light are used to provide lighting in lieu of the sun. Other factors that may be considered are; ambient temperature, nutrient solution temperature, photoperiod, and air humidity [8].

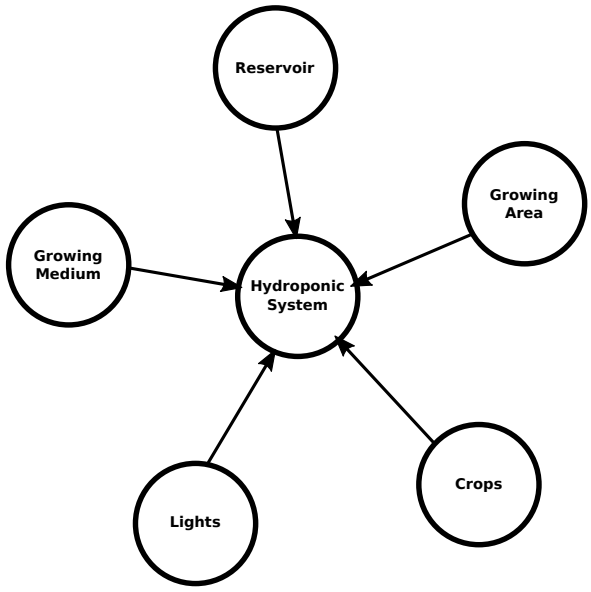


Figure 1: Components of a hydroponic system

[*]Corresponding Author Adam Adam, Department of Computer Science, Yobe State University, (+234) 806 067 3838 & adam@ysu.edu.sa

1.1 Types of Hydroponics

Figure 2 shows a basic hydroponic system – all hydroponic systems can be developed by modifying it. Hydroponic techniques are divided into seven types, Wick, Deep Water Culture (DWC), Ebb and Flow (Flood and Drain), Drip (recovery or non-recovery), Nutrient Film Technique (NFT), Aeroponics [11] and Fogponics.

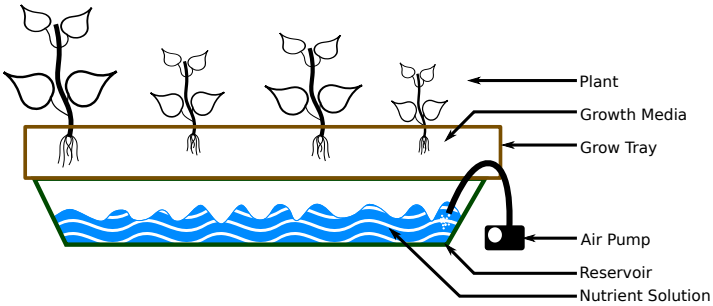


Figure 2: A Hydroponic system

1.1.1 Wick Hydroponic

As the name implies, the wick type hydroponics system feeds the plants with the nutrient solution via a wick. The most widely used wicks are Pro-Mix, Vermiculite, Perlite and Coconut Fiber [12]. The system is easy to maintain because it is a passive system without any moving parts not even an air pump since the plants' roots are not submerged in the nutrient solution. However, its biggest weakness is that the wick can only deliver a small quantity of water to the plants at a time. Hence, bigger plants may be starved.

1.1.2 Deep Water Culture (DWC) Hydroponics

DWC is also known as water culture hydroponics. In this system, the growth medium is made up of Styrofoam, which floats directly on the nutrient solution. Oxygen delivery is the difficult part of this system [13]. Since the plants' roots are completely submerged in the nutrients a system of air pump and air stone provides them with oxygen. DWC is commonly used in plants the need a large quantity of water to grow, such as lettuce.

1.1.3 EBB and Flow (Flood and Drain) Hydroponics

In this system, a submersible pump is installed in the nutrient solution where it pumps the solution up and into the growth tray to flood it. A system is installed such that it allows the solution to ebb back into the reservoir. The pump is controlled with a timer such that it turns on the pump to fill the growth tray and then turn it off so that the solution flows back slowly into the reservoir. This flood-ebb cycle allows the root to get a solution (during the flood cycle) as well as oxygen (during the ebb cycle). However, plants' root are susceptible to diseases [14]. Since the plants' life entirely depend on the flood-ebb cycle, then by extension it means they depend on the timer, the pump and the drain system, which significantly reduces the reliability of the system.

1.1.4 Drip Hydroponic

Here the system uses drip to feed the plants with the nutrients from the reservoir. Excess nutrients are either fed-back into the reservoir or it is allowed to drain away or evaporate. The earlier is known as recovery drain hydroponics while the latter is called a non-recovery drain hydroponic system [15]. The advantage of this system is that it can be tailored to any type of plant since the flow rate of the nutrient can be adjusted. However, it is difficult to maintain because of the pH shift in the recovery type.

1.1.5 Nutrient Film Technique (NFT)

NFT pumps nutrient solution to a tilted growth tray. The nutrient flow to the other end of the tray where it is drained back to the reservoir. The system requires no growth medium and timer. Furthermore, the air pump is not needed since the growth medium of the system is air. Unfortunately, the plants are prone to wilting when the flow of nutrient flow stops because the roots dry relatively fast [16].

1.1.6 Aeroponics

This system gets rid of the barrier between the growth tray and the reservoir. The roots are left dangling in the reservoir above the solution. A mist of the solution is sprayed to the roots at regular intervals. Aeration is the primary advantage of this system. Also, a smaller amount of nutrient is consumed compared to other systems [17]. However, the system needs a shorter time interval for the nutrient cycle. This is tantamount to more energy consumption. Moreover, the plants' root would quickly dry should the pump or timer fails.

1.1.7 Fogponics

A more advanced variant of aeroponics is Fogponics. Here too, the growth medium is air. However, fog emitter (also known as fogger) is used to produce smaller droplets (ranging $5 - 30\ \mu m$) than in aeroponics. The fog produced delivers water and nutrients to the plants' root. Fogponics is better than aeroponics because; smaller (fog) droplets encourage more nutrient absorption [18], the fog can also reach more parts of the plants' root compared to spray droplets and the lack of flowing nutrient solution allows for easier crop monitoring [19].

1.2 Motivation

A lot of work has been carried out to improve and/or automate hydroponic systems: Sambo et. al [20], presents a literature survey of hydroponic systems and smart technologies that can be used to improve the performance of such systems. Kaewwiset and Yooyativong [9] investigated the relationship between the electrical conductivity (EC) and the pH of the hydroponic nutrient mixing system. In [21], potassium acrylate is proposed for hydroponic substrates, because of its water-absorbent nature and the fact that it is biodegradable. In another research [22], the best method to reduce power consumption while maintaining optimum growth is investigated.

The authors found that Far Red (FR) treatment leads to late flowering and slow stem growth. The impact of microalgae on hydroponics is investigated in [23]. The authors argue that they can be used to purify wastewater which can be used in hydroponics.

In this paper, the latest effort of researchers to automate the hydroponic system using embedded systems and the Internet of Things (IoT) is reported. In addition, a nomenclature of a smart hydroponic system base on the studied literature is presented. The remaining part of the paper is as follows: Section 2 presents an analysis of the smart hydroponic systems in the literature and developed a nomenclature for them. Section 3 lists the challenges in smart hydroponics and recommends some possible research areas. Section 4 concludes the paper.

2 Smart Hydroponic Systems

Automated hydroponic systems have been investigated in the literature for a long time. This is because the automated systems have the advantage of providing huge throughput, high efficiency and require no specialized manpower. Furthermore, the automation of hydroponic is not beyond reach since the tasks carried out in a hydroponic system can easily be automated. Also, computers have the capacity to do it better, because of its high availability compare to humans.

In this paper, the smart hydroponic systems are classified into four (4) main categories. The categories are developed based on the Automation level, Tasks automated, Type of Automation and Mode of control:

2.1 Automation Level

This section explores the classification of hydroponic systems based on the level of automation. Hydroponics can be classified as; 1) semi-automated, and 2) fully automated. The earlier is a group of hydroponics systems that control only part of the system shown in Figure 1, while the latter is a group that covers all hydroponics system where all its components are automatically controlled.

2.1.1 Semi-automated

Savvas [24], proposed two alternative techniques for automatic nutrient replenishment in closed-loop hydroponic systems. Both proposed systems use electric conductivity and pH as input and addition of nutrients or freshwater as output. The first system uses several pumps (one for each nutrient and another for freshwater) while the second has only two pumps (one for fertilizer and the other for freshwater). The author concluded that both systems can be used in replenishing nutrients in hydroponic systems. Also, Domingues et al. [25], developed an Automated Hydroponic System (AHS) that uses pH, temperature and electric conductivity as parameters for measuring the nutrients level. They argue that temperature is a very important parameter since it is directly related to both pH and electric conductivity. The system is used in growing lettuce (Lactuca sativa L.). The proposed AHS uses a system of electric valves, one for each nutrient. They are used to control the pH level of the nutrient solution. The authors show that the proposed system is as good as hydroponics system control by experts.

Phutthisathian et. al [26] developed an Ontology-based nutrient solution control system for hydroponics was developed. The authors argue that building a smart hydroponic system is a complex task and it should be designed using Ontology Web Language (OWL) and Resource Description Framework (RDF). The system controls the nutrient supply through two nutrient tanks and pH levels through the use of a separate acid tank. The system also has a fourth tank for water supply. The system uses EC to monitor the nutrient solution. The authors build an ontology of control variables from OWL and RDF. Ontology is able to design structure and mechanism that makes hydroponic systems more effective.

2.1.2 Fully Automated

Yamaguchi et. al [27] developed a small closed hydroponic system capable of cultivating 10 heads of lettuce is developed in this paper. In this paper, a closed hydroponic system is a hydroponics system designed to be used indoors or in a glasshouse, where lighting could be controlled. The system has sensors for monitoring the temperature and humidity of the system and its outside surroundings. It also has a light sensor for measuring the ambient light. Also, the system has white LEDs, a cooling fan, an air pump to supply oxygen to the nutrient solution. A Raspberry Pi with a camera is also installed to observe the system. The authors also used renewable energy to manage the hydroponic system. A hybrid energy control system using lead-acid battery and hydrogen fuel cell battery was deployed. The proposed system has the potential of bringing hydroponic farming to every home, because of its ease of setup. Furthermore, the use of renewable energy reduces the carbon footprint of the system. The authors were able to harvest the lettuce after 28 days. Also, the coexistence of the system and occupants prove fruitful due to the abundant CO_2 provided by the later. The proposed system requires human intervention in order to run efficiently.

Aquaponics is a type of hydroponic system that uses aquaculture (fish farming) as the main source of nutrients from the fish wastes. The plants remove the nitrogen and phosphorous introduced in the water by the fish thus cleaning the water. In this paper [28], an automated closed (i.e. indoor) aquaponics is developed. Managed by an Arduino Mega microcontroller, the system feeds the fish, monitors and controls the water temperature and level. The pump is used to irrigate the Water spinach (Ipomoea aquatica) and Spinach (Spinacia oleracea) used in the system. In addition, the stand-alone system uses an SD card to keep a log of the activities of the microcontroller. However, the system depends on fixed parameters set by the user, also the system uses a 100W light source which may increase the temperature of the aquaponic system.

2.2 Automated Task

Smart hydroponic systems can be classified based on the task that is automated. Therefore, smart hydroponics can be designed to offer assistance to the hydroponic system by maintaining the system's hardware, providing consumables or securing (cyber or physical) the system. Another category helps in crop cultivation by providing the plant with nutrient, lighting or proving cultivation related crop maintenance, like moving seedlings from nursery to hydroponics system, replacing dead plants and harvesting. The final subcategory

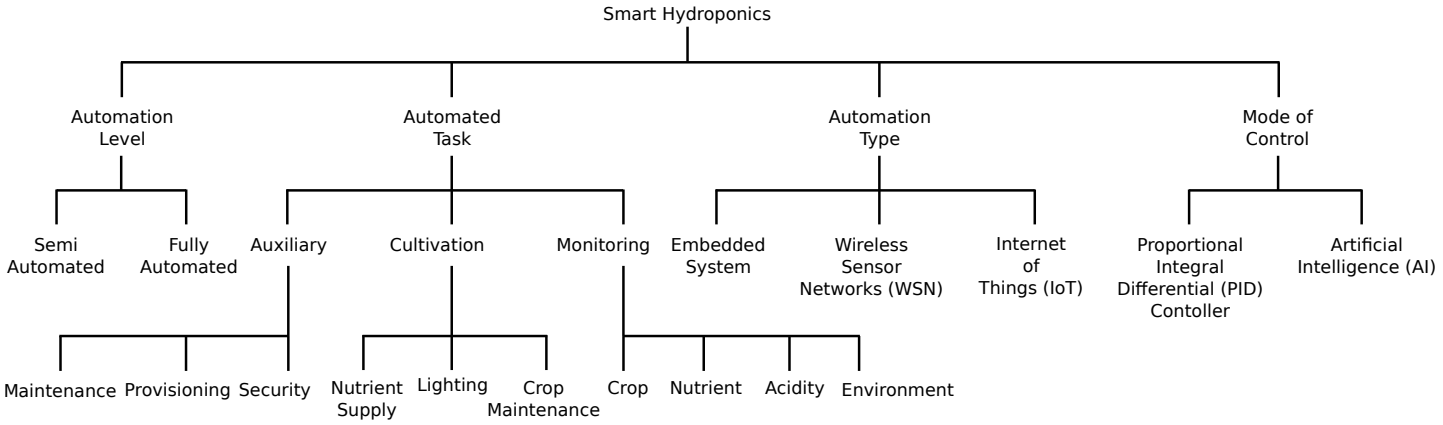


Figure 3: Taxonomy for Smart Hydroponic Systems

monitors some or all aspects of the hydroponic system and reports it to the farmer.

2.2.1 Auxiliary

Several smart systems that assist hydroponic systems in crop cultivation have also been proposed in the literature. For example, Tu et al. [29] developed a water harvesting device called Cactus. It is designed to extract water from the air. It can be used in conjunction with IoT to develop a hydroponic system in areas with little or no water. Cactus enable farmers to grow crops in inhospitable areas around the world. It can use solar panels or wind turbines as sources of energy because the system requires a small amount of energy. However, farmers have to account for the latency of the system, since it takes time for the system to condense water from the atmosphere. Therefore, they have to provide enough water to the hydroponic system before the device starts generating its own water.

Saenz et. al [30] developed strawberries collecting a robot for greenhouse hydroponic systems. The proposed system uses artificial vision for identifying the stage of ripeness based on the redness of the strawberries. The system is developed using OpenCV libraries and Arduino Mega for the image processing (sensor), while PUMA manipulator simulated using Solidworks and MatLab for robotic arm (Actuator). The authors focus their automation to harvesting crops rather than their cultivation. The proposed system automates the harvesting of crops, which helps reduce labour cost and the introduction of pests and diseases often brought in by workers into the hydroponic system. The system has shown good performance. It can also be used in harvesting other crops. The system's use of Arduino microcontrollers for image processing may cause an increase in the latency of the proposed system.

Takeuchi [31, 32, 33] pointed out how 3D printed component can be used to develop a hydroponic system and by so doing reducing the cost of hydroponic culture. The paper investigated the use of 3D hydroponic substrate using SBS, PLA, ABS and flexible TPU plastic against the traditional rock wool and sponge. The authors developed custom software that can be used to specify the plant-covered region (in 2D) on the planting substrate. This enables the software to locate places where seeds will be placed. The authors found that the SBS foundation (substrate) holds relatively modest water compare to the traditional hydroponic substrates. The PLA, ABS and flexible TPU perform worse because they do not give way for the root to grow while SBS does. The printed SBS material does not have good water retention (like rock wool), therefore it does not guarantee germination of seeds. This problem is solved by automatically coating the plant seeds in superabsorbent polymer gel (sodium polyacrylate) before placing them in the planting substrate. Due to hardware limitations, not all seeds can be planted, since different seeds have different sizes and shapes. Furthermore, the system has low accuracy in seed placement. Moreover, once the plants germinate they must be irrigated manually.

Sunlight is the most important part of a greenhouse, but dust that accumulates on the house makes it difficult for light to pass through. Cleaning the greenhouse is therefore necessary. Currently, this is done manually, which consumes time and is dangerous for the workers. As such, [34] developed a cleaning root from wood and aluminum frame and a soft-sponge curved wiper blades driven by 12v DC motor. This makes the robot light-weight and easy to use. The robot is controlled by an MCS-51 microcontroller. Experiments show that the proposed system cleans $85 - 95\%$ of the dust on the house at a speed of 1.3 meters per minute. This is an improvement to cleaning systems in the literature who can only work on flats surfaces which is impractical for greenhouses. The system is light-weight therefore less energy is needed to operate it. The paper concludes that the proposed system is low cost and can be used on plastic sheet greenhouses. The system is open-loop, it contains no feedback that indicates the cleanliness of the surface. Therefore and the operator has to be available on-site to overlook it.

Some tasks are difficult to automate using sensors and actuators. Therefore [35] proposed the use of a robot to remotely manage a hydroponic farm. The proposed system uses Microsoft's Kinect v2 for controlling the robot using hand gestures. This allows farmers to manage their farms without any computer training. The proposed system can be used in kick-starting vegetation in remote areas. The authors found that users can carry out activities like lifting, placing, and grasping. However, placing objects in the targeted area is found to be difficult compared to other activities.

2.2.2 Cultivation

The majority of hydroponic systems are developed to help manage the cultivation of crops. Eridani et. al [36] proposed automation of Nutrient Film Technique (NFT) hydroponic system using Arduino Uno as a controller, Total Dissolved Solids (TDS) Sensor for testing nutrient level, a GP2Y0A21 proximity sensor as a water level detector, and servo for replenishing the nutrient from a container. The authors use the electric conductivity (EC) sensor as a TDS sensor. However, it was calibrated using a TDS meter. The author derived a polynomial relating the reading from the TDS meter to the TDS sensor values. The use of the TDS sensor is cheaper than the TDS meter. The proposed system supply more nutrients when the solution is below 800 ppm. Also, the TDS sensor is up to 97.8% accurate. The system controls the amount of nutrient solution in the tank rather than the concentration. Hence, the system may supply more nutrients than the plants needed.

A DIY sensor-based automatic control mobile application for hydroponics is developed in [37, 38]. It is a cheaper and more affordable way of developing small scale hydroponic systems. The system monitors the ambient temperature, water temperature, humidity, light intensity, pH and EC. The system then uses a set of rules to control the water chiller, LED, light shade, fog spray, and fan in order to manage the environment of the hydroponics. Status of the system is stored in the cloud and the users are allowed to monitor the system through a mobile application. The mobile application also allows farmers to record their yield and a reminder to notify farmers on harvesting date. In a similar experiment is developed by [39]. However, there are some modifications: 1) The hydroponics system is enclosed in a greenhouse, 2) sprinklers are used to control the ambient temperature of the hydroponic system but not the solution, and 3) there are no LEDs to control the lighting of the system. The authors recorded poor performance compared to traditional farming. This shows that the factors selected by [37] matter more than those selected in [39].

Dbritto et. al [40] developed an automated hydroponic system for growing Tomato F1 Hybrid Suhyana seeds. The system is an indoor hydroponic system where amount of water, light, nutrients, and temperature are controlled using Raspberry Pi 3 model B. The authors classify readings from the sensor as follows; Temperature Levels (Cold, Warm, Hot), Water Levels (Low, Normal, High), pH levels (Acid, Neutral, Alkaline) and Light Intensity Levels (Low, Normal, High). They argue that categorizing the reading will make it easier for researchers to develop a model to automate a system whose output sill simply be appropriate or inappropriate. The paper concluded that their taxonomy system can be used for an AI-controlled hydroponic system. However, classification of data to universal categories may be difficult since each plant has a different requirement which may not necessarily fall in the range of other plants.

Arif et al. [41], argue that plants require different nutrients as they age. Therefore an automated hydroponic system needs to be able to determine the age of a plant in order to ensure accurate nutrient supply. The authors in [42] investigated the performance of 3 classical machine learning classifiers: decision tree, Naive Bayes, Multi-Layer Perceptron; and one type of deep neural network in the detecting the freshness of vegetation harvested from a hydroponic system. The proposed system uses image processing and machine learning technologies to detect fresh and withered vegetables. The experiment shows that the decision tree (J48) model was found to have the best accuracy of 98.12%. This system can be used in harvesting and/or monitoring the health of crops grown in a hydroponic system. Also, [41] develop a system that determines the age of plants from their picture using digital image processing technique. The proposed system uses Raspberry Pi 3, a camera module to acquire and process that information which is then saved in a web server. The system is deployed in an NFT (Nutrient FilmTechnique) hydroponic lettuce farm. The authors used vegetation cover to deduce the age of the plant. Vegetation cover is obtain using pixel count after masking to isolate the plant. The proposed system requires neither training nor dataset. The proposed system is able to predict the age of plants with an accuracy of 80%. However, the correlation between the age of a plant and its vegetation cover must be known.

Namgyel et. al [43] developed a smart hydroponic system with LED lighting technology enabled by the IoT system uses the Wi-Fi module as a mode of data transmission and monitoring. SHT 15 sensor was used to measure air temperature and relative humidity. In order to measure and monitor the nutrient temperature, a waterproof temperature sensor is confined in the nutrient solution tank. ML8511 sensors were used to measure the solar energy and surveillance camera to keep track of the physiological growth of the plants. Germination of lettuce seed was done before the experiment and then the seedlings were moved to the hydroponic system 7 days after. The architecture of the system is then divided into four segments where the first three were irradiated with deep red LEDs (640 – 660 nm), deep royal blue LEDs (440 – 450 nm) and combination of deep red LEDs (640 – 660*nm*) and deep royal blue LEDs (440 – 450 nm) with the ratio of (1:1) respectively. Solar radiation was the sole light source in the fourth growing segment. Weatherboard embedded with the system controller unit in order to read and receive data from different sensors and transmit to node MCU with embedded ESP8266 Wi-Fi microprocessor and the data flow from sensors are then uploaded to the server through local Wi-Fi network. The system has the ability to monitor and make data accessible to users in real-time. Results indicate that blue supplementary LED light fostered positive effects on lettuce growth, morphology and pigment content.

2.2.3 Monitoring

Tagle et. al [44] developed a data acquisition system for hydroponics farming. A DFRobot Arduino Mega 2560 was used as a microcontroller. The system monitors vital parameters such as air temperature, relative humidity, water temperature, water level, pH level, and light intensity. Also, the acquired data is used to troubleshoot the system in times of failure. An Arduino data logger was designed to collect data from the sensors after every 5 minutes and then store them in an SD card. A 4X20 LCD module was integrated to allow real-time access to the system's performance. The data gathered by all sensors which were stored in an SD card showed fluctuation in environmental condition occurring in the system in real-time which assured ease in monitoring and areas with problems were easily identified for improvement. The data acquisition system

showed the performance of the hydroponic tower with respect to the growth factors. After validation, it is found that the average percentage error of each growing condition based on their respective standard instrument are near 5% which shows that sensors available in the market can be used to accurately monitor a hydroponic system.

Recently, [45] successfully developed a hydroponic system that measures only the temperature of the nutrient solution and water level. The system can be monitored through a smartphone. Authors in [46, 47], developed a system that measures water level, pH, electrical conductivity, water temperature, and relative humidity of the system. The data is sent to a database and farmers are notified of events such as the imbalance of nutrients in the system. The system is only designed to monitor the hydroponic environment. Therefore, the system has no actuators and rely on the farmer to manage events [46]. While [47] can control the pH level, as well as supply water it harvested from rain to the system.

However, there is more to monitoring pH using pH sensor than meets the eye. Many pH meters need re-calibration as they age. A solution is proposed in [48]. The authors proposed an auto-calibration system for monitoring and controlling the pH (in addition to temperature and EC). The system consists of a chamber where the pH meter is placed. A series of pumps are connected to the chamber. Three of the pumps are for calibration where one pumps acid, the other pumps base and the third pump buffer (liquid with pH 7). The other pump samples nutrient solution from the hydroponic reservoir to the pH meter's chamber to testing. The last pump pumps water to the chamber to ensure that the pH meter is clean between the calibration and the testing cycles and increases water level in the hydroponic reservoir. The system was monitored for fifteen (15) days with the aim to maintain a pH from 5.5 to 5.7. A mean pH of 5.65 and standard deviation of 0.058 was recorded when using the system, while a mean pH of 5.00 and a standard deviation of 0.36 is observed on the control system. Although the system shows a promising performance, the large number of moving parts involved will reduce its reliability.

Social media applications have also been used in monitoring hydroponics. Sisyanto et. al [49] developed hydroponic smart farming that can be monitored online via telegram messenger. The system uses different sensors to automatically monitor the vital parameters and environment of the hydroponic system an post it on Telegram messenger using Raspberry Pi 3. The system allows collaboration between hydroponic farmers. With telegram messenger hydroponic farmers communicate with the group/each other and with the telebot one can request the condition of the environment in a hydroponic system and reply back to the sender.

2.3 Automation Type

The type of automation is also very important. In this paper, a category is given for the different types of automation. A smart hydroponic system can be a purely embedded system where automation and data remain in-situ. Users have to be physically on the site to monitor or adjust the settings of the system. Another subcategory is Wireless Sensor Networks (WSN) where a distributed network of an embedded system is controlled by a single base station (also known as gateway) users connected to the gateway can access the information. Finally, the IoT based hydroponic system is basically WSNs connected to the internet to improve the scalability and availability of information to the user/farmer.

2.3.1 Embedded System based Hydroponic System

This category presents embedded systems that are mainly made up of electronic sensors, controllers and actuators. Usually, data is saved in a memory device where the user can copy it. Settings and configurations can only be done in-situ. For example, in [50, 51], only pH was measured. The authors develop an AHS using Arduino Mega 2560 micro-controller and ATLAS and ATC pH sensors [51]. Users would key in the pH level for the targeted crop and the system will maintain that pH for the targeted hydroponic system. The system was tested for five (5) days with an average deviation of 0.58 and 0.93 for ATLAS and ATC respectively. However, pH is only one factor that affects plant growth, therefore, optimal plant growth is not possible. The authors used Deep Water Culture (DWC) which is a hydroponic technique where water containing nutrients are fed directly to the roots of the plant continuously. The system consists of a pH and water level sensor, an actuator system, using a valve controlled solution mixer and water pump [50]. The proposed system uses a glass electrode pH sensor which measures both pH and temperature and feeds it to the microcontroller SK 40C for decision making. The setup ensures reduced energy consumption. The authors establish that there is a need for at least 5.64 ml of acid and alkaline to increase the acidity and alkalinity by 0.312 pH and 0.244 pH respectively. These levels can be achieved through the use of a feeder valve. The glass electrode meter is prone to errors [52].

More sensors were added in [53, 54], a sensor module capable of measuring EC, water temperature and water level of a hydroponic system. The system uses the capacitance of a cable ribbon to measure the level of the water. It uses a TI TMP20 sensor coated by resin connected at the end of the aforementioned ribbon cable to measure the temperature and electrodes to measure the EC of the nutrient solution. An oscillator is used to measure the capacitance of the cable. The proposed sensor module shows that capacitance and water levels are proportional, which is easy for sensing and analysis. The authors found that the accuracy of the proposed system is within 1% of commercial sensors. However, they pointed out that the structure of the electrodes will be reviewed because the error in EC increases when the distance between the electrodes changes. The sensor module shows a logarithmic relationship with the capacitance which is difficult to convert to digital at high levels of EC since the change is small.

Velazquez et. al [55] proposed an automated hydroponic system for a cherry tomato. The system ensures that the pH is maintained between 5.5 to 6 and the electric conductivity (salinity) between 1.8 mS to 3 mS. The system uses PIC16F877A microcontroller and an ion-selective electrode (ISE) is used for pH measurement while a pair of gold wires separated 1 cm apart is used for EC sensor. The authors added conditioning circuits and amplifiers to the pH sensor and the EC sensor. The use of amplifiers and conditioning circuit help reduce the amount of noise in the readings. The results showed high accuracy on pH measurements with a correlation coefficient of 0.994 for expected values and a standard deviation of 0.01 after repeated measurements. The use of extra circuitry will increase

energy consumption and latency of the proposed system.

A semi-automatic hydroponic system that controls the humidity, ambient temperature and pH is developed in [56]. The system switches between automatic monitoring and manual control. This allows the system to maintain optimal conditions for the plant regardless of whether the user is connected or not. Users/farmers can also monitor the system through an LCD screen on site. For more flexibility, Nalwade and Mote [57] developed an automated hydroponic system that ensures the potential of Hydrogen (pH) and Electric Conductivity (EC) of nutrients is maintained by the system and measured automatically by sending the information to the farmer through GSM. Aeration problem is solved using an oxygen pump. The energy problem solved using solar panels. Similarly, [58] proposed a cost-effective and environmentally sound automated solar-powered aquaponics system, for local communities. It consists of a water circulation system that circulates water to an aquaculture tank and hydroponic beds, aquaponics control and monitoring system actuators, GSM shield and NI LabVIEW, solar energy conversion system, cooling, and heating systems that maintain the air and water temperature. However, the nutrient level from the fish tank is not monitored. This may lead to a high concentration of nutrients supplied to the plant.

2.3.2 *Wireless Sensor Networks (WSN)*

Wireless Sensor Network (WSN) is the network of computing devices with limited resources (such as processor, memory, energy) aimed at sensing their environment [59, 60, 61]. WSN finds application in health-care, military, manufacturing industry and agriculture [62, 63, 64, 65].

Figure 4 shows a typical WSN. It consists of: Sensor nodes who are responsible for sensing the environment, Relay nodes are responsible for forwarding information to the sink node. Relay nodes are needed because nodes in WSN are short-range devices. Sink nodes connect the sensor network with the outside world – usually a workstation. Actuator nodes are responsible for manipulating the environment; such as turning on/off valves, finally, the user (workstation) is responsible for collecting and analyzing data from the network. In this research, WSN will be used to automate the hydroponic system in order to enable hydroponic farming easier and cost-effective.

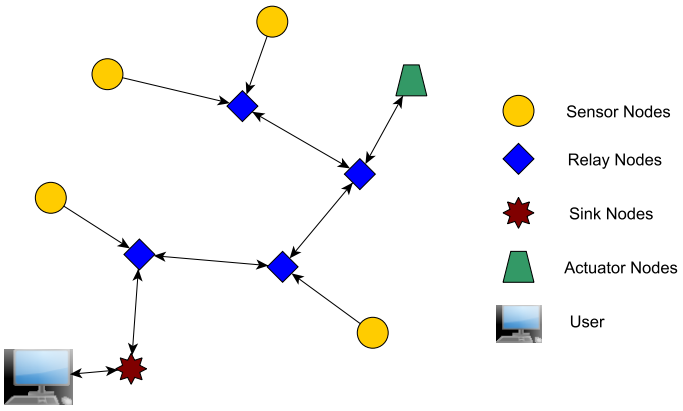


Figure 4: A typical Wireless Sensor Network (WSN).

Jai et. al [66] proposed a system consisting of soil moisture, pH level, light intensity, water temperature, humidity and temperature sensor. The sensors forward their data to the control unit which is made up of Arduino Uno microcontroller. A predefined threshold is set for the moisture sensor. When the level is below a certain threshold the microcontroller increases the amount of water circulating in the system. The soil moisture sensors are placed in every pot of the system to ensure proper nutrient supply. Since all nodes are individually monitored to ensure adequate water and nutrient supply. The system ensures no water wastage and that the plants got adequate water and nutrient supply. The system uses a predetermined threshold and the water level is compensated proportionally. This leads to oscillation in the system.

When using WSN for hydroponics there is need to choose a suitable transmission protocol that will ensure low energy consumption and low latency in communication. The aforementioned factors are the most important because most smart hydroponic systems transmit only commands and state of the system. As such, researchers use zigbee transmission protocol. In [67], zigbee stack is used to develop a real-time smart hydroponic system. The authors found that although the performance of the system is satisfactory, the distance of the system from a the control station is inversely-proportionally affecting the performance of the system. In addition, the authors cite size of memory of the sensor node as a challenge to the efficiency of the system.

2.3.3 *Internet of Things (IoT) based Hydroponic System*

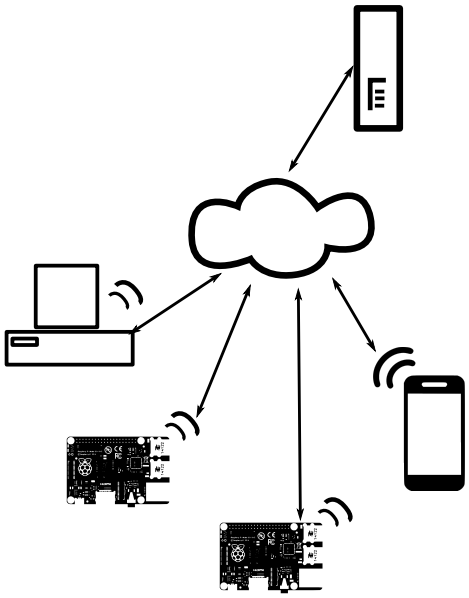


Figure 5: A typical Internet of Things (IoT) setup.

As shown in Figure 5, IoT is a network of embedded systems, smartphones and personal computers that communicate through the internet. Data generated from IoT systems are saved and monitored through a server or a broker depending on the design. Researches have been carried out on the use of IoT in hydroponic systems [68]. IoT based hydroponics system is designed to monitor and/or control a hydroponic system with the help of a server. The system commu-

nicates with the server through Ethernet, Wi-Fi, GPRS or satellite radio communication.

Fernandes et al. [69] developed an Arduino and Raspberry Pi-based solution which are connected with USB cable and Pi card memory to save the experimental data. The system was able to allow external users and programmers to connect to the Pi via the internet. The system was able to control the temperature and humidity of the greenhouse based hydroponic system. The system improves the efficiency of the growth of crops. The system was proven to be much faster compared with traditional techniques. However, the configuration of the system was not easy due to the extra plug-ins and third-party software required. Also, the system allows the monitoring of plant growth by applying Image Processing. unfortunately, the system was designed to favor crop production in the summertime only, since it provides only cooling systems to control High temperature. Furthermore, neither the pH or Nutrient solution concentration (EC) levels were measured so, the information obtained from the system regarding nutrient combination and pH will not be reliable.

Siregar et al. [70] proposed a closed hydroponic monitoring system designed with PH sensor, Electric Conductivity Sensor, water temperature sensor, air temperature, Light Sensor, GSM / GPRS, Open Garden Shield, Open Garden Hydroponic, and Arduino Uno microcontroller. The system was used to test lettuce, red spinach, and mustard Pak Choy plants. All sensors in the proposed system are combined using Open Garden Shield and Open Garden Hydroponics. The open garden shield is an open-source Arduino shield that is used in automating gardens. The data received by the shield from the sensors is sent to the Arduino Uno using I2C communication. Finally, the data is sent to a remote server with the help of a GSM shield. The proposed system used the open garden shield and open garden hydroponics which makes the reproduction of results easy. The authors conclude that the lettuce, red spinach, and mustard Pak Choy plants perform well. The proposed system has no actuators to control the environment, it depends on the farmer for control.

The authors in [71] developed an automated aquaponic system with minimum resource requirements. The system automatically feeds the fish and water the plants with the help of the GSM network. The proposed system consists of two (2) modules; Automatic water supply and Automatic fish food feeder. The earlier module is responsible for supplying the plants with nutrient water from the aquarium while monitoring the water level of the aquarium. Whenever the level of water in the aquarium goes down, the module replenishes it from a secondary tank. The second module is responsible for feeding the fish. Furthermore, the system also has heater and temperature sensors for controlling the temperature of the water in the fish tank. It also controls the lighting provided by Light Emitting Diodes (LEDs) to ensure photosynthesis. The proposed system is portable, therefore it can be installed at home or office. Furthermore, the system is made up of a few components therefore is it cheap to develop. The authors observe that the proposed system consumes only 3.87 L of water in two weeks. Also, the fish consume $5-10$ gm of food in two days. It is also found that an automated aquaponic system with an aquarium and plant bed is feasible. The proposed system lacks acidity and basicity control system. Also, no system is provided to ensure the well being of the fish.

In [72] developed a small scale intelligent plant care hydroponic box using IoTtalk [73]. The system allows users to easily add, remove or exchange sensors and actuators in the plant-care system and it also provides environment driven control methods between the sensors and actuators. In order to make analysis, the sensors generate environment data and send it to a server through Arduino Yun via wi-fi, then data processing system executes the environment driven controls function based on the user-defined threshold values and finally the IoTtalk plant-care intelligence directs the actuators in the response system to take action. Similar IoTtalk-based hydroponic systems have been proposed in [74, 75]. ThingSpeak, like IoTtalk is an IoT platform [76]. Lakshmiprabha et. al [77] used ThingSpeak to manage and monitor water level, humidity and temperature with the help of Rasberry Pi 3 B+ and Field-Programmable Gate Array (FPGA). Also, a larger model of the system is developed by in [78, 79]. The system allows for modularity, thereby enabling ease of maintainability. However, the system relies on communication with a server that may not be available all the time. Also, a fixed threshold may not be optimal since the needs and consumption rate of the plants' changes as they grow.

An IoT based smart hydroponic system is presented in [80] where the sensor communicates with the server through the Ethernet. The authors developed a closed hydroponic control and monitoring system with the help of Arduino Uno, a clock module (RTC I2C), a relative humidity and temperature sensor (DHT11), a water pump (HJ 701), a fan (POLUX PH 2009) and an electrical resistance and an aerator (AC 500) was also used to oxygenate the water. An Ethernet module is used to send readings of the different sensors to a server. The system is developed for the Stevia plant. Also, aeration and lighting are accurately controlled with the help of the RTC clock. The proposed system is cheaper than the most commercially available hydroponic system. The authors of the proposed system were able to save to 76% of the cost price of a commercially available AeroFlo hydroponic system.

IoT based hydroponic systems are known to have many interconnected devices, as well as software. As such [81, 82] proposed a cloud-based IoT for controlling and monitoring hydroponic systems using publish-subscribe middleware over MQTT protocol. The use of middleware allows the network administrator to set the quality of service for the system which will help guarantee delivery even if the network itself is not reliable. The authors in IoT based hydroponic systems are known to have many interconnected devices, as well as software. As such [81] found that the response time of the system is 7.21 seconds. Also, they found that the system needs at least 21.55 seconds to determine if a node is offline. Authors in [83] integrated TLS/SSL to encrypt MQTT packets. Moreover, the authors used Raspberry Pi as the broker/server, where the Ardunio (client) sends its data including the image of the plants. The remote management system was developed using open source libraries.

To solve the volume and velocity problem of data developed in a hydroponic system, data fusion is used in [84]. Data fusion is a technique where information from different sensors is combined (fuse) to achieve a more accurate result. Data fusion also reduces the transmissions in the network thereby reducing traffic and energy consumption. The authors proposed an automatic control and management system for tropical hydroponic cultivation. The system monitors the power status, temperature, humidity, water

level, pH, and EC data which are sent to a remote database over the internet. Users can access the status of the hydroponic sensor through a mobile application. The system is a two-tier network containing sensors (S_i) and data fusion sensors (DF_i). The sensors forward their sensed data to the data fusion sensors where the data is combined to achieve a more accurate result before saving it in the remote database. The authors used data fusion techniques to improve the accuracy of data received from the sensors. This is achieved by fusing information from the ultrasonic sensor with a solenoid valve to get water level and SHT31 sensor with a foggy solenoid valve for temperature and humidity control. The system also notifies the user of a power outage. The system is able to improve the accuracy of readings from the sensor nodes without either taking several readings or using many different sensors.

Another solution is the use of edge computing. Edge computing is an IoT paradigm where processing devices (with communication capabilities) are placed at the edge of the network close to the user, to provide services to the user in lieu of the server [85]. The author in [86] argue that edge computing is the cost effective way of ensuring connectivity in smart hydroponics. In their paper, the authors proposed a three-tier system with local, edge and cloud layer. The local layer is responsible for monitoring the crops and carrying out atomic action on them at real-time. The edge layer is responsible for monitoring and managing the hydroponic system, while the cloud is responsible for data collection and analysis. The system uses FIWARE, ORION Context Broker [87] over MQTT to allow the edge and the cloud layers to communication efficiently. The edge layer uses Network functions virtualization (NFV), which increases it flexibility. The system reduce water consumption by 30% and some nutrients by 80%.

2.4 Mode of Control

In smart hydroponic systems where actuators are involved, controlling how the actuators behave is a very important aspect of the system. In this paper, mode of control is categorized into Proportional Integral and Derivative (PID) controller or Artificial Intelligence (AI). The PID controllers are the type of controllers developed using control engineering techniques they are widely used because they are flexible and easy to develop [88]. AI can also be used to control how the actuators perform their actions to ensure they do not damage the system or themselves. In this paper, a broader interpretation of the word AI is used; any algorithm that can perform human tasks is placed in this category.

2.4.1 Proportional Integral Derivative (PID) Controller

Proportional, Integral and Derivative (PID) controller is a control system that uses feedback to adjust the output to the desired level [89]. PID controllers are popular because they are easy to develop and tunning, wide availability and high success rest.

Asumadu et. al [90] proposed a computer-based hydroponic control system using LABVIEW. The system measures the pH level using EC and the water level using an ultrasonic sensor. The control system adjusts the water level in the system by observing the error between the current level and the predetermined level. The use of a computer and LABVIEW improves the precision of the height. Also, the experiment shows that a hydroponic setup can be controlled through an instrumentation software.

However, since the system is a proportional controller pH level will oscillate. The work carried out in [91] is a testament to that fact. The authors developed an automated hydroponic system for green mustered. The system ensures that the plants are kept in a deep water culture (DWC) of pH between 6.0-7.5, which is set by a keypad and managed by Arduino Mega 2560 microcontroller. The authors pointed out that there is a delay in the pH level change after more fertilizer is added. Therefore, they used two tanks (main tank and hydroponic tank), the main tank is used to verify the pH level of the water before is it finally pumped to the hydroponic tank where is it used for growing the plants. The use of a separate tank allows the authors to stir the liquid and observe its response to the added nutrients. The authors found that after a certain time (16 days) the system begins to oscillate. They also pointed out that the reading accuracy of the pH meter used has a major role in the performance of the system.

To reduce oscillations encountered in a proportional controller, Adhau et. al [92] developed an automated closed hydroponic system with proportional integral and derivative (PID) controller is developed with the help of AVR microcontroller and LabVIEW. The system uses DHT11 humidity sensor, LM 325 temperature sensor, electrodes for pH, and ambient light sensors, while it uses fans, blowers, halogen light bulbs, pumps and motors as actuators to control the state of the system. The authors used a PID controller which allows their system to control the environmental parameters with little or no oscillation. The PID is tuned using Zeigler-Nicholas closed-loop tuning methods. The authors also used the AVR microcontroller rather than the traditional LabIEW's data acquisition (DAQ) card, which reduces the cost of implementing the system. The proposed system improves the response time of the controller with minimal oscillation. The system shows great stability and smooth control actions due to the use of a PID controller. The requirements of a hydroponic system change as the plants grow, however, the proposed system uses fixed values which require human interference in order to perform efficiently.

2.4.2 Artificial Intelligence (AI) based Hydroponic System

Gartphol et. al [93] developed predictive models for Lettuce quality from the Internet of Things-based hydroponic farms. The authors developed a smart hydroponic lettuce farm using the Internet of Things to gather environmental data and control the farm's operation in real-time. A large dataset was generated at the end of the experiment which was used to create regression models using machine learning techniques. Some of the features used were environmental data such as the amount and intensity of light, humidity, temperature, together with weekly measurements of plant growth. Internet of Things allows automatic operation which can be accessed within a household or office area. A good learning result has not been achieved by the model due to errors such as those found in measuring the lettuce's height because of changes caused by blowing wind which causes the crop to sway. Other errors are those found from the stem's width. In [94], Bayesian Network is reported to be 84.53% accuracy and a 66.67% more yielded crops than the manual system.

Fuzzy logic has been used to solve a variety of problems in an

automated hydroponic system [95, 96]. Mashumah et. al [95] developed a Nutrient Film Technique (NFT) hydroponics that uses fuzzy logic control to set the threshold value for the electrical conductivity (EC) level. The system uses a webcam Logitech B525 camera to capture the image of a plant and converts it to HSV Histogram to provide EC setpoint value. An ultrasonic distance sensor HC-SR04 measures the volume of the water with an error rate of 15.6%. The saturation value of the leaves is used to determine the set point of EC value for the control of the system. The saturation value is chosen because of its high level of linearity with the age of the plant. EC sensor SKU:DRF0300 measures the concentration of the nutrient in the water. It then uses EC error and water volume as inputs of fuzzy logic to adjust the opening time of the nutrients and water valves. The system can provide and retain the EC value given according to the age of the plant with an error rate of 8.9%. In another research, scientists investigate the most suitable colour space for vegetation image processing in smart hydroponics [97]. The authors [97] show that the HSV method has an average accuracy of 75.33% while the CIELAB method has a greater average accuracy of 78.39%. Research conducted in [98, 96] shows that the steady-state time (i.e. time to reach the targeted value) of fuzzy logic-based EC controller is around 110 sec, which is quite slow.

Also, researchers in [99] developed a novel fuzzy logic-based control system combined with a genetic algorithm (GA) to maintain the pH level of a hydroponic system. The authors argue that a fuzzy logic control (FLC) system is more suitable than PID controllers because PID controllers suffer limited feedback. The system has five storage tanks: acid, base, buffer, nutrient A and nutrient B. The tanks are controlled by valves in order to maintain the pH level. A drain valve is used to control the level of mixture in the hydroponic tank. The nutrient solution is monitored using a pH meter and EC meter. The authors used 199 fuzzy rules obtained from experimental observations and expert opinions from researchers in MCRC, India. The authors developed a novel Mamdani fuzzy inference system (FIS), which uses a given set of control parameters to grade the quality of a hydroponic solution. The result is then fed to a GA as a fitness function in order to obtain the optimal valve for the control signals. The authors pointed out that their proposed FIS performs better than fuzzy or PID controllers. The proposed algorithm proved efficient in terms of convergence rate and resource utilization compared to conventional error based algorithms used in solution control. However, the use of a drain valve in the proposed system will lead to wastage.

The authors in [100] grew wheat (Triticum aestivum L.) in glass housed hydroponic environment in order to classify nutrient deficiency using multilayer perceptron (MLP) neural network with using hyperspectral data as input. The authors create a nutrient mixture with the investigated nutrient absent. Then spectral reflectance of the leaves in wavelengths 401 nm to 770 nm is fed to the 3-layer (input, hidden, and output) MLP. The aim of the research is to investigate the performance of MLP in classifying leaves with a deficiency in deficient in nitrogen (N), phosphorus (P), potassium (K), and calcium (Ca). This research can be used to automate the fertilizer application process in hydroponic systems. The authors show that MLP with backpropagation algorithm used for classification performs (N 90.9%, P 100%, K 90%, and Ca 100%) better than regression model (N 93.0%, P 87.2%, K 91.9%, and Ca 97.4%). Although the Artificial Neural Network (ANN) shows excellent performance, training data set for other crops and vegetables may not be easy to come by.

Furthermore, [101] proposed an ANN technique for controlling the light intensity by observing the net photosynthesis of the plants in a hydroponic system. The net photosynthesis is measured using the Koito model KMC series assimilation chamber and C02 analyzer, while the light intensity is measured using the LI-COR model LI-185B quantum sensor. The light intensity and net photosynthesis were measured at an interval of 2 minutes from 8 o'clock to 11 o'clock. This data is then fed it to the neural network for training. The ANN uses backpropagation for training. The proposed system is self-adaptive since it can learn from any given environment. The authors found that there is no significant difference between the 3-layer and 4-layer ANN. Also, the authors compared a 3-layer ANN, 3-layer recurrent ANN, Group Method of Data Handling (GMDH), and least mean square techniques. It is found that there is no significant difference in the performance of the first three techniques, while the last-mentioned technique is performed the worse of them. Single feature (light intensity) was used in measuring the performance of the proposed system which is not enough to account for the net photosynthesis of the hydroponic system.

More dimension was added to the ANN in [102]. Here, the authors developed a hybrid 4-stage control system using neural networks and genetic algorithm. The system's objective is to maximize the total leaf length to stem diameter (TLL/SD), which is an indicator of a healthy plant. The neural network is used to identify the environmental factors that affect the plant's physiology (i.e. TLL/SD). Then the genetic algorithm is used to control them. The 3-layer neural networks take nutrient concentration (NC) and light intensity as input and TLL-SD ratio as output. It was trained using the backpropagation technique. The neural network is used to obtain fitness functions for the nutrient concentration at 4 stages. These fitness functions at then used to obtain the nutrient concentrations at the four stages (NC_1, NC_2, NC_3 and NC_4). The proposed system uses two AI techniques to observe the environment and to detect the right nutrient concentration for the plant at different stages of its life. Therefore, the system is adaptive to the environment and the type of plant. The proposed system shows that the NC should be a little high in the first stage of the plants' growth, then a little low in the second stage, then a little higher than the first stage in the third stage and a little high than the third stage in the fourth stage. The system was observed for 24 days and the TLL-SD ratio shows an increase of ≈ 9.375 per day. The proposed system needs a dataset which is hard to come by.

3 Challenges in Automating Hydroponic Systems

So far, researchers have made a tremendous contribution to the development of efficient smart hydroponic systems. However, several challenges in this area are still untouched. Some of these challenges are:

1. Security of the system is a very important problem. For companies and countries to cozy up to the smart hydroponic

Table 1: Summary of the Automated Hydroponic Systems

Papers	Automation Level		Automated Task											Automated Type			Mode of Control		Comment
			Auxiliary				Cultivation			Monitoring									
	Semi Automated	Fully Automated	Maintenance	Provisioning	Harvesting	Security	Nutrient Supply	Light Supply	Crop Maintenance	Crop	Nutrient	Environment	Acidity	Embedded System	WSN	IoT	PID	AI	
[24, 25, 26, 50, 51, 57, 91]	✓						✓				✓		✓	✓			P		
[27, 40]		✓					✓	✓		✓	✓	✓	✓	✓					Mainly for house hold
[28]		✓		✓			✓	✓			✓	✓	✓	✓					Aquaponics provides nutrients to the hydroponics system.
[29]				✓															Provide the system with water by condensing it from the atmosphere
[30]					✓														
[31, 32, 33]				✓															3D printed growth medium
[34]			✓											✓					Cleans the green house.
[35]			✓	✓	✓	✓	✓	✓	✓	✓		✓							Robotics.
[36]	✓						✓				✓			✓			P		Only monitors level.
[37, 38]		✓					✓			✓	✓	✓	✓			✓		✓	System is rule based. Also, Temperature of Solution, Humidity, pH, Nutrients and light were controlled.
[39]		✓					✓				✓	✓	✓			✓	P		Humidity, pH and Nutrients but no light were controlled.
[41]	✓						✓			✓	✓		✓			✓	P		
[42]					✓					✓				✓					
[43]	✓					✓	✓	✓		✓	✓	✓	✓			✓			Survaillence camera is used to monitor plants

Table 2: Summary of the Automated Hydroponic Systems Continued

Papers	Automation Level		Automated Task											Automated Type			Mode of Control		Comment
			Auxiliary				Cultivation			Monitoring									
	Semi Automated	Fully Automated	Maintenance	Provisioning	Harvesting	Security	Nutrient Supply	Light Supply	Crop Maintenance	Crop	Nutrient	Environment	Acidity	Embedded System	WSN	IoT	PID	AI	
[44]			✓								✓	✓	✓	✓					used for troubleshooting hydroponic system
[45]	✓						✓				✓	✓				✓	P		
[46]											✓	✓	✓			✓			
[47]	✓			✓			✓			✓	✓	✓	✓	✓			P		
[48]	✓											✓		✓		✓			
[49]										✓	✓	✓	✓			✓			System uses social media
[53, 54, 56]	✓						✓				✓	✓	✓	✓			P		
[55]											✓		✓	✓					
[58]		✓		✓			✓	✓				✓			✓				Nutrient is provided by the aquaponic system
[66]	✓						✓			✓	✓	✓	✓		✓				
[67]	✓						✓				✓	✓	✓	✓	✓				
[69]	✓						✓			✓						✓	P		Temperature control system
[70]											✓	✓	✓			✓			
[71, 72, 74, 75, 78, 79]		✓					✓				✓	✓	✓			✓	P		
[77, 84]	✓						✓				✓	✓	✓			✓	P		
[80]		✓					✓	✓			✓	✓	✓	✓		✓			
[81, 82, **?**]		✓					✓				✓	✓	✓			✓	P		Middleware is used to help with heterogeneity
[83]		✓				✓	✓			✓	✓	✓	✓			✓	P		
[86]		✓					✓	✓			✓	✓	✓	✓		✓			Edge computing is used to reduce latency
[90]	✓						✓				✓		✓				P		Computer is used for automation
[91]	✓						✓				✓	✓		✓			P		
[92]		✓					✓	✓			✓	✓	✓		✓		PID		
[93, 94]		✓					✓			✓	✓	✓	✓			✓		✓	
[95]	✓						✓			✓	✓		✓		✓				
[98, 96]	✓						✓				✓		✓		✓			✓	Fuzzy logic only

Table 3: Summary of the Automated Hydroponic Systems Continued

Papers	Automation Level		Automated Task											Automated Type			Mode of Control		Comment
			Auxiliary				Cultivation			Monitoring									
	Semi Automated	Fully Automated	Maintenance	Provisioning	Harvesting	Security	Nutrient Supply	Light Supply	Crop Maintenance	Crop	Nutrient	Environment	Acidity	Embedded System	WSN	IoT	PID	AI	
[99]	✓						✓				✓		✓		✓			✓	Fuzzy logic combined with Genetic algorithm to control pH
[100, 101]	✓							✓				✓		✓				✓	Computer directly reads the sensors. ANN is used for AI.
[102]	✓						✓	✓			✓	✓						✓	ANN is used for AI.

system they need to be certain that their system is secured from intruders.

2. Automation of harvesting in hydroponics with the help of IoT has not been explored. Since plants are grown in batches, the manual harvesting method may lead to the infection of the younger crops with human beings as vectors. Furthermore, since the automated system can work tirelessly, it is more economical to use smart systems in the harvesting of the crops.

3. Impact of the sensors used in the development of plants has not been investigated. For example, what is the effect of electricity on plants?

4. Further investigation should be made into the tolerance plants to high EC and/or pH. The knowledge of how tolerant plants are at extreme levels of pH will allow system designers to develop systems that can optimally manage the hydroponic systems.

5. The use of controllers should also be investigated. Stress on fish (in aquaponics) and plants due to fluctuations in EC and pH should also be investigated. Cheaper controllers are prone to oscillation, therefore it is necessary to investigate how far a fluctuation hydroponic systems and aquaponic systems are willing to tolerate.

6. AI-based systems are difficult to develop due to the lack of availability of datasets. Therefore, there is a need for more datasets to allow researchers to develop efficient AI-based hydroponic systems.

4 Conclusion

In this paper smart hydroponic systems in the literature have been investigated. Table 1, 2 and 3 summarize the survey carried out in this paper. It is noted that researchers have done a lot of work in the monitoring of hydroponic systems and the automation of nutrient supply. However, the automation of nutrient supply is mainly controlled by proportional controllers, who are notorious for oscillations. The few AI-based research carried out have shown that AI is a good technique for controlling the hydroponic system, but they are expensive and require huge computation. Also, it is noted that IoT is the most attractive form of automation type, because of the flexibility it offers during system design. This is all the more reason why the security of the hydroponic system must be developed.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank the Tertiary Education Trust Fund (TETFund), Yobe State University (YSU) Damaturu Nigeria and King Fahd University of Petroleum and Minerals (KFUPM) for their support in this research.

References

[1] T. Baras, *DIY Hydroponic Gardens: How to Design and Build an Inexpensive System for Growing Plants in Water*. Cool Springs Press, 2018. [Online]. Available: https://books.google.com.sa/books?id=rwlMDwAAQBAJ

[2] D. Singh, J. Davidson, and M. Books, *Introduction to Hydroponics - Growing Your Plants Without Any Soil*, ser. Gardening Series. Mendon Cottage Books, 2016. [Online]. Available: https://books.google.com.sa/books?id=RAMtDQAAQBAJ

[3] M. Raviv and J. Lieth, *Soilless Culture: Theory and Practice*. Elsevier Science, 2007. [Online]. Available: https://books.google.com.sa/books?id=NvDHJxRwsgYC

[4] W. Ke and Z. Xiong, "Difference of growth, copper accumulation and mineral element uptake in two elsholtzia haichowensis populations under copper and mineral nutrition stress," in *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*. IEEE, 2008, pp. 4704–4708.

[5] H. Wang, Y. Wang, and Y. Yang, "Effects of exogenous phenolic acids on roots of poplar hydroponic cuttings," in *2011 International Conference on Remote Sensing, Environment and Transportation Engineering*. IEEE, 2011, pp. 8245–8249.

[6] N. Suzui, N. Kawachi, M. Yamaguchi, N. S. Ishioka, and S. Fujimaki, "A monitoring system of radioactive tracers in hydroponic solution for research on plant physiology," in *2009 1st International Conference on Advancements in Nuclear Instrumentation, Measurement Methods and their Applications*. IEEE, 2009, pp. 1–3.

[7] M. Liu, X. Xi, S. Wang, Y. Xu, and W. Song, "Research on differences of component and quantity of organic acids in the root exudates among the three green manures," in *World Automation Congress 2012*. IEEE, 2012, pp. 1–4.

[8] G. Gupta, *Text Book of Plant Diseases*. Discovery Publishing House, 2004. [Online]. Available: https://books.google.com.sa/books?id=OuoicDXQ-xYC

[9] T. Kaewwiset and T. Yooyativong, "Estimation of electrical conductivity and ph in hydroponic nutrient mixing system using linear regression algorithm," in *2017 International Conference on Digital Arts, Media and Technology (ICDAMT)*. IEEE, 2017, pp. 1–5.

[10] J. Muro, I. Irigoyen, P. Samitier, P. Mazuela, M. Salas, J. Soler, and M. Urrestarazu, "Wood fiber as growing medium in hydroponic crop," in *International Symposium on Soilless Culture and Hydroponics 697*, 2004, pp. 179–185.

[11] A.-K. Ammar, A. Anas, D. Ehab, and H. Yazan, "Hydroponics system," pp. 14–16, 2019. [Online]. Available: https://www.researchgate.net/publication/330578599_Hydroponics_Integrated_with_PV_System

[12] M. Honeydew, *Gardening*. Lulu.com, 2016. [Online]. Available: https://books.google.com.sa/books?id=4EMyDAAAQBAJ

[13] J. D. Butcher, C. P. Laubscher, and J. C. Coetzee, "A study of oxygenation techniques and the chlorophyll responses of pelargonium tomentosum grown in deep water culture hydroponics," *HortScience*, vol. 52, no. 7, pp. 952–957, 2017.

[14] J. Jones, *Hydroponics: A Practical Guide for the Soilless Grower*. CRC Press, 2016. [Online]. Available: https://books.google.com.ng/books?id=y_bKBQAAQBAJ

[15] N. JOSHI, *GREEN SPACES: CREATE YOUR OWN*. Notion Press, 2018. [Online]. Available: https://books.google.com.sa/books?id=CA9tDwAAQBAJ

[16] G. Van Patten, *Hydroponic Basics*. Van Patten Publishing, 2004. [Online]. Available: https://books.google.com.sa/books?id=L87bl7EIYkkC

[17] M. Calvin, *Foundations of Space Biology and Medicine. Volume 3: Space Medicine and Biotechnology*. National Aeronautics and Space Administration, 1975. [Online]. Available: https://books.google.com.sa/books?id=iOA_AQAAMAAJ

[18] M. C. Watson, "Fogponic plant growth system," Dec. 27 2018, uS Patent App. 15/974,356.

[19] P. Wootton-Beard, "Growing without soil: An overview of hydroponics," *Farming Connect*, 2019. [Online]. Available: https://pure.aber.ac.uk/portal/files/30769801/technical_article_hydroponics_final.pdf

[20] P. Sambo, C. Nicoletto, A. Giro, Y. Pii, F. Valentinuzzi, T. Mimmo, P. Lugli, G. Orzes, F. Mazzetto, S. Astolfi *et al.*, "Hydroponic solutions for soilless production systems: Issues and opportunities in a smart agriculture perspective," *Frontiers in plant science*, vol. 10, 2019.

[21] O. Adrianes and G. M. S. Zarazúa, "Potassium acrylate: A novelty in hydroponic substrates," in *2017 XIII International Engineering Congress (CONIIN)*. IEEE, 2017, pp. 1–8.

[22] A. N. Harun, R. Ahmad, and N. Mohamed, "Plant growth optimization using variable intensity and far red led treatment in indoor farming," in *2015 International Conference on Smart Sensors and Application (ICSSA)*. IEEE, 2015, pp. 92–97.

[23] S. Huo, J. Liu, M. Addy, P. Chen, D. Necas, P. Cheng, K. Li, H. Chai, Y. Liu, and R. Ruan, "The influence of microalgae on vegetable production and nutrient removal in greenhouse hydroponics," *Journal of Cleaner Production*, vol. 243, p. 118563, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095965261933433X

[24] D. Savvas, "SW–soil and water: Automated replenishment of recycled greenhouse effluents with individual nutrients in hydroponics by means of two alternative models," *Biosystems Engineering*, vol. 83, no. 2, pp. 225 – 236, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1537511002901066

[25] D. S. Domingues, H. W. Takahashi, C. A. Camara, and S. L. Nixdorf, "Automated system developed to control ph and concentration of nutrient solution evaluated in hydroponic lettuce production," *Computers and electronics in agriculture*, vol. 84, pp. 53–61, 2012.

[26] A. Phutthisathian, N. Pantasen, and N. Maneerat, "Ontology-based nutrient solution control system for hydroponics," in *2011 First International Conference on Instrumentation, Measurement, Computer, Communication and Control*. IEEE, 2011, pp. 258–261.

[27] S. YAMAGUCHI, Y. TAKAHASHI, and T. HAYASHI, "Small indoor hydroponic system with renewable energy," in *2018 18th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2018, pp. 313–318.

[28] M. F. Saaid, N. Fadhil, M. M. Ali, and M. Noor, "Automated indoor aquaponic cultivation technique," in *2013 IEEE 3rd International Conference on System Engineering and Technology*. IEEE, 2013, pp. 285–289.

[29] D. T. C. Tu, P. B. Duong, N. T. H. Vuong, P. H. Luan, N. H. Nghia, P. N. Dien, N. H. Quang, and C. H. Buu, "The cactus device - an ultimate solution for irrigation," in *2016 3rd International Conference on Green Technology and Sustainable Development (GTSD)*, Nov 2016, pp. 270–273.

[30] E. Saenz, M. Jimenez, and A. Ramirez, "Strawberries collecting robot prototype in greenhouse hydroponic systems," in *Symposium of Signals, Images and Artificial Vision-2013: STSIVA-2013*. IEEE, 2013, pp. 1–4.

[31] Y. Takeuchi, "3d printable hydroponics: A digital fabrication pipeline for soilless plant cultivation," *IEEE Access*, vol. 7, pp. 35 863–35 873, 2019.

[32] Y. Takeuchi, "Printable hydroponic gardens: Initial explorations and considerations," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '16. New York, NY, USA: ACM, 2016, pp. 449–458. [Online]. Available: http://doi.acm.org/10.1145/2851581.2892587

[33] ——, "Printable hydroponics: Digital fabrication of ecological systems," in *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, ser. ISS '18. New York, NY, USA: ACM, 2018, pp. 433–435. [Online]. Available: http://doi.acm.org/10.1145/3279778.3279929

[34] N. Seemuang, "A cleaning robot for greenhouse roofs," in *2017 2nd International Conference on Control and Robotics Engineering (ICCRE)*. IEEE, 2017, pp. 49–52.

[35] M. Sekimoto, K. Ikeshiro, and H. Imamura, "Development of interface using marker less ar for hydroponic culture managing systems in the distant place," in *2017 IEEE 12th International Conference on Power Electronics and Drive Systems (PEDS)*. IEEE, 2017, pp. 843–847.

[36] D. Eridani, O. Wardhani, and E. D. Widianto, "Designing and implementing the arduino-based nutrition feeding automation system of a prototype scaled nutrient film technique (nft) hydroponics using total dissolved solids (tds) sensor," in *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. IEEE, 2017, pp. 170–175.

[37] C. Peuchpanngarm, P. Srinitiworawong, W. Samerjai, and T. Sunetnanta, "Diy sensor-based automatic control mobile application for hydroponics," in *2016 Fifth ICT International Student Project Conference (ICT-ISPC)*. IEEE, 2016, pp. 57–60.

[38] N. K. Jacob, "Iot powered portable aquaponics system," in *Proceedings of the Second International Conference on Internet of Things, Data and Cloud Computing*, ser. ICC '17. New York, NY, USA: ACM, 2017, pp. 66:1–66:5. [Online]. Available: http://doi.acm.org/10.1145/3018896.3018965

[39] T. Changmai, S. Gertphol, and P. Chulak, "Smart hydroponic lettuce farm using internet of things," in *2018 10th International Conference on Knowledge and Smart Technology (KST)*. IEEE, 2018, pp. 231–236.

[40] G. Dbritto and S. Hamdare, "An ai based system design to develop and monitor a hydroponic farm," in *2018 International Conference on Smart City and Emerging Technology (ICSCET)*, Jan 2018, pp. 1–5.

[41] A. Nursyahid, M. R. Wibisono, E. D. Wardihani, T. A. Setyawan *et al.*, "Plant age identification system of outdoor hydroponic cultivation based on digital image processing," in *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. IEEE, 2017, pp. 213–218.

[42] K. Wongpatikaseree, N. Hnoohom, and S. Yuenyong, "Machine learning methods for assessing freshness in hydroponic produce," in *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*. IEEE, 2018, pp. 1–4.

[43] T. Namgyel, S. Siyang, C. Khunarak, T. Pobkrut, J. Norbu, T. Chaiyasit, and T. Kerdcharoen, "Iot based hydroponic system with supplementary led light for smart home farming of lettuce," in *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. IEEE, 2018, pp. 221–224.

[44] S. Tagle, R. Pena, F. Oblea, H. Benoza, N. Ledesma, J. Gonzaga, and L. A. G. Lim, "Development of an automated data acquisition system for hydroponic farming," in *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*, Nov 2018, pp. 1–5.

[45] P. Sihombing, N. A. Karina, J. T. Tarigan, and M. I. Syarif, "Automated hydroponics nutrition plants systems using arduino uno microcontroller based on android," *Journal of Physics: Conference Series*, vol. 978, p. 012014, mar 2018. [Online]. Available: https://doi.org/10.1088%2F1742-6596%2F978%2F1%2F012014

[46] P. Belhekar, A. Thakare, P. Budhe, U. Shinde, and V. Waghmode, "Automated system for farming with hydroponic style," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Aug 2018, pp. 1–4.

[47] F. Jaimes, B. Collazos, E. Arce, and M. Chauca, "Hydroponic system with automated hydrolysis using renewable energy self-sustainable," in *MATEC Web of Conferences*, vol. 256. EDP Sciences, 2019, p. 02012.

[48] C. Cambra, S. Sendra, J. Lloret, and R. Lacuesta, "Smart system for bicarbonate control in irrigation for hydroponic precision farming," *Sensors*, vol. 18, no. 5, p. 1333, 2018.

[49] R. E. N. Sisyanto, N. B. Kurniawan *et al.*, "Hydroponic smart farming using cyber physical social system with telegram messenger," in *2017 International Conference on Information Technology Systems and Innovation (ICITSI)*. IEEE, 2017, pp. 239–245.

[50] M. Saaid, N. Yahya, M. Noor, and M. M. Ali, "A development of an automatic microcontroller system for deep water culture (dwc)," in *2013 IEEE 9th International Colloquium on Signal Processing and its Applications*. IEEE, 2013, pp. 328–332.

[51] M. F. Saaid, A. Sanuddin, M. Ali, and M. S. A. I. M. Yassin, "Automated ph controller system for hydroponic cultivation," in *2015 IEEE Symposium on Computer Applications Industrial Electronics (ISCAIE)*, April 2015, pp. 186–190.

[52] W. Martindale and P. S. of Great Britain, *The extra pharmocopoeia*, ser. The extra pharmocopoeia. Pharmaceutical Press, 1955, no. v. 2. [Online]. Available: https://books.google.com.sa/books?id=8cxQAAAAYAAJ

[53] T. Nishimura, Y. Okuyama, and A. Satoh, "High-accuracy and low-cost sensor module for hydroponic culture system," in *2016 IEEE 5th Global Conference on Consumer Electronics*. IEEE, 2016, pp. 1–4.

[54] T. Nishimura, Y. Okuyama, A. Matsushita, H. Ikeda, and A. Satoh, "A compact hardware design of a sensor module for hydroponics," in *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*. IEEE, 2017, pp. 1–4.

[55] L. Velazquez, M. Hernandez, M. Leon, R. B. Domínguez, and J. Gutierrez, "First advances on the development of a hydroponic system for cherry tomato culture," in *2013 10th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*. IEEE, 2013, pp. 155–159.

[56] S. Charumathi, R. Kaviya, J. Kumariyarasi, R. Manisha, and P. Dhivya, "Optimization and control of hydroponics agriculture using iot," *Asian J. Appl. Sci. Technol*, vol. 1, no. 2, pp. 96–98, 2017.

[57] R. Nalwade and T. Mote, "Hydroponics farming," in *2017 International Conference on Trends in Electronics and Informatics (ICEI)*. IEEE, 2017, pp. 645–650.

[58] A. M. Nagayo, C. Mendoza, E. Vega, R. K. Al Izki, and R. S. Jamisola, "An automated solar-powered aquaponics system towards agricultural sustainability in the sultanate of oman," in *2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC)*. IEEE, 2017, pp. 42–49.

[59] F. Aliyu, S. umar, and H. Al-Duwaish, "A survey of applications of artificial neural networks in wireless sensor networks," 2019, the 8th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO'2019).

[60] F. Aliyu and T. Sheltami, "Development of an energy-harvesting toxic and combustible gas sensor for oil and gas industries," *Sensors and Actuators B: Chemical*, vol. 231, pp. 265 – 275, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925400516303318

[61] F. Aliyu, M. Al-shaboti, Y. Garba, T. Sheltami, A. Barnawi, and M. A. Morsy, "Hydrogen sulfide (h2s) gas safety system for oil drilling sites using wireless sensor network," *Procedia Computer Science*, vol. 63, pp. 499 – 504, 2015, the 6th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2015)/ The 5th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2015)/ Affiliated Workshops. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050915025107

[62] J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Terzis, and M. Welsh, "Wireless sensor networks for healthcare," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1947–1960, 2010.

[63] M. A. Hussain, K. kyung Sup *et al.*, "Wsn research activities for military application," in *2009 11th International Conference on Advanced Communication Technology*, vol. 1. IEEE, 2009, pp. 271–274.

[64] Y. Wang, X. Yin, and D. You, "Application of wireless sensor networks in smart grid," *power System technology*, vol. 34, no. 5, pp. 7–11, 2010.

[65] R. Hussain, J. Sahgal, P. Mishra, and B. Sharma, "Application of wsn in rural development, agriculture water management," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 2, no. 5, pp. 68–72, 2012.

[66] N. Jai, B. Dontha, A. Tripathy, and S. Mande, "Near real time-sensing system for hydroponics based urban farming," in *2018 3rd International Conference for Convergence in Technology (I2CT)*. IEEE, 2018, pp. 1–5.

[67] K. Kalovrektis, C. Lykas, I. Fountas, A. Gkotsinas, I. Lekakis *et al.*, "Development and application embedded systems and wireless network of sensors to control of hydroponic greenhouses," *International Journal of Agriculture and Forestry*, vol. 3, no. 5, pp. 198–202, 2013.

[68] R. Vidhya and K. Valarmathi, "Survey on automatic monitoring of hydroponics farms using iot," in *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2018, pp. 125–128.

[69] M. B. Fernandes, B. A. Costa, and J. M. Lemos, "Hydroponic greenhouse crop optimization," in *2018 13th APCA International Conference on Control and Soft Computing (CONTROLO)*. IEEE, 2018, pp. 270–275.

[70] B. Siregar, S. Efendi, H. Pranoto, R. Ginting, U. Andayani, and F. Fahmi, "Remote monitoring system for hydroponic planting media," in *2017 International Conference on ICT For Smart Society (ICISS)*. IEEE, 2017, pp. 1–6.

[71] K. S. Aishwarya, M. Harish, S. Prathibhashree, and K. Panimozhi, "Survey on iot based automated aquaponics gardening approaches," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, April 2018, pp. 1495–1500.

[72] T.-H. Wu, C.-H. Chang, Y.-W. Lin, L.-D. Van, and Y.-B. Lin, "Intelligent plant care hydroponic box using iottalk," in *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, 2016, pp. 398–401.

[73] Y.-B. Lin, Y.-W. Lin, C.-M. Huang, C.-Y. Chih, and P. Lin, "Iottalk: A management platform for reconfigurable sensor devices," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1552–1562, 2017.

[74] W.-L. Chen, Y.-B. Lin, Y.-W. Lin, R. Chen, J.-K. Liao, F.-L. Ng, Y.-Y. Chan, Y.-C. Liu, C.-C. Wang, C.-H. Chiu *et al.*, "Agritalk: Iot for precision soil farming of turmeric cultivation," *IEEE Internet of Things Journal*, 2019.

[75] L.-D. Van, Y.-B. Lin, T.-H. Wu, Y.-W. Lin, S.-R. Peng, L.-H. Kao, and C.-H. Chang, "Planttalk: A smartphone-based intelligent hydroponic plant box," *Sensors*, vol. 19, no. 8, p. 1763, 2019.

[76] M. A. G. Maureira, D. Oldenhof, and L. Teernstra, "Thingspeak–an api and web service for the internet of things," *World Wide Web*, 2011.

[77] K. E. Lakshmiprabha and C. Govindaraju, "Hydroponic-based smart irrigation system using internet of things," *International Journal of Communication Systems*, p. e4071, 2019.

[78] P. N. Crisnapati, I. N. K. Wardana, I. K. A. A. Aryanto, and A. Hermawan, "Hommons: Hydroponic management and monitoring system for an iot based nft farm using web technology," in *2017 5th International Conference on Cyber and IT Service Management (CITSM)*, Aug 2017, pp. 1–6.

[79] P. N. Crisnapati, I. N. K. Wardana, I. K. A. A. Aryanto, and A. Hermawan, "Hommons: Hydroponic management and monitoring system for an iot based nft farm using web technology," in *2017 5th International Conference on Cyber and IT Service Management (CITSM)*. IEEE, 2017, pp. 1–6.

[80] G. J. U. Ossa, C. G. Montoya, and M. A. V. Velásquez, "Control and monitoring of a stevia plantation in a hydroponic system," in *2017 IEEE 3rd Colombian Conference on Automatic Control (CCAC)*, Oct 2017, pp. 1–5.

[81] M. A. Triawan, H. Hindersah, D. Yolanda, and F. Hadiatna, "Internet of things using publish and subscribe method cloud-based application to nft-based hydroponic system," in *2016 6th International Conference on System Engineering and Technology (ICSET)*. IEEE, 2016, pp. 98–104.

[82] S. Ruengittinun, S. Phongsamsuan, and P. Sureeratanakorn, "Applied internet of thing for smart hydroponic farming ecosystem (hfe)," in *2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media)*, Aug 2017, pp. 1–4.

[83] A. Satoh, "A hydroponic planter system to enable an urban agriculture service industry," in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*. IEEE, 2018, pp. 281–284.

[84] J. Chaiwongsai, "Automatic control and management system for tropical hydroponic cultivation," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019, pp. 1–4.

[85] F. Aliyu, T. Sheltami, and E. M. Shakshuki, "A detection and prevention technique for man in the middle attack in fog computing," *Procedia Computer Science*, vol. 141, pp. 24–31, 2018.

[86] M. A. Zamora-Izquierdo, J. Santa, J. A. Martínez, V. Martínez, and A. F. Skarmeta, "Smart farming iot platform based on edge and cloud computing," *Biosystems Engineering*, vol. 177, pp. 4 – 17, 2019, intelligent Systems for Environmental Applications. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1537511018301211

[87] I. Batyrshin, M. de Lourdes Martínez-Villaseñor, and H. Espinosa, *Advances in Computational Intelligence: 17th Mexican International Conference on Artificial Intelligence, MICAI 2018, Guadalajara, Mexico, October 22–27, 2018, Proceedings*, ser. Lecture Notes in Computer Science. Springer International Publishing, 2019, no. pt. 2. [Online]. Available: https://books.google.com.ng/books?id=PUOCDwAAQBAJ

[88] I. Boiko, *Non-parametric Tuning of PID Controllers: A Modified Relay-Feedback-Test Approach*, ser. Advances in Industrial Control. Springer London, 2012. [Online]. Available: https://books.google.com.sa/books?id=vE6PHWgnRgsC

[89] M. Chidambaram and N. Saxena, *Relay Tuning of PID Controllers: For Unstable MIMO Processes*, ser. Advances in Industrial Control. Springer Singapore, 2018. [Online]. Available: https://books.google.com.sa/books?id=GGRLDwAAQBAJ

[90] J. A. Asumadu, B. Smith, N. S. Dogan, P. A. Loretan, and H. Aglan, "Microprocessor-based instrument for hydroponic growth chambers used in ecological life support systems," in *Quality Measurement: The Indispensable Bridge between Theory and Reality (No Measurements? No Science! Joint Conference-1996: IEEE Instrumentation and Measurement Technology Conference and IMEKO Tec*, vol. 1. IEEE, 1996, pp. 325–329.

[91] M. F. Saaid, A. Sanuddin, M. Ali, and M. S. A. I. M. Yassin, "Automated pH controller system for hydroponic cultivation," in *2015 IEEE Symposium on Computer Applications Industrial Electronics (ISCAIE)*, April 2015, pp. 186–190.

[92] S. Adhau, R. Surwase, and K. H. Kowdiki, "Design of fully automated low cost hydroponic system using labview and avr microcontroller," in *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, March 2017, pp. 1–4.

[93] S. Gertphol, P. Chulaka, and T. Changmai, "Predictive models for lettuce quality from internet of things-based hydroponic farm," in *2018 22nd International Computer Science and Engineering Conference (ICSEC)*. IEEE, 2018, pp. 1–5.

[94] M. I. Alipio, A. E. M. D. Cruz, J. D. A. Doria, and R. M. S. Fruto, "On the design of nutrient film technique hydroponics farm for smart agriculture," *Engineering in Agriculture, Environment and Food*, vol. 12, no. 3, pp. 315 – 324, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1881836617303294

[95] S. Mashumah, M. Rivai, and A. N. Irfansyah, "Nutrient film technique based hydroponic system using fuzzy logic control," in *2018 International Seminar on Intelligent Technology and Its Applications (ISITIA)*. IEEE, 2018, pp. 387–390.

[96] D. Yolanda, L. Arief, T. A. Sundara, M. Deza, and D. Oktavia, "Control of electrical conductivity for nft hydroponic systems using fuzzy logic and android environment," in *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*. IEEE, 2018, pp. 508–514.

[97] T. A. Setyawan, S. A. Riwinanto, Suhendro, Helmy, A. Nursyahid, and A. S. Nugroho, "Comparison of hsv and lab color spaces for hydroponic monitoring system," in *2018 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Sep. 2018, pp. 347–352.

[98] D. Yolanda, H. Hindersah, F. Hadiatna, and M. A. Triawan, "Implementation of real-time fuzzy logic control for nft-based hydroponic system on internet of things environment," in *2016 6th International Conference on System Engineering and Technology (ICSET)*. IEEE, 2016, pp. 153–159.

[99] L. M. JSM and C. Sridevi, "Design of efficient hydroponic nutrient solution control system using soft computing based solution grading," in *2014 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)*. IEEE, 2014, pp. 148–154.

[100] T. Ayala-Silva and C. A. Beyl, "Use of neural networks to discriminate between control leaves of wheat or those deficient in nitrogen, phosphorus, potassium, and calcium using spectral data," in *Proceedings of the 5th Biannual World Automation Congress*, vol. 13, June 2002, pp. 49–57.

[101] T. Morimoto, I. Cho, and Y. Hashimoto, "Identification of hydroponics in an advanced control system of the greenhouse," *IFAC Proceedings Volumes*, vol. 25, no. 15, pp. 169 – 174, 1992, 9th IFAC/IFORS Symposium on Identification and System Parameter Estimation 1991 , Budapest, Hungary, 8-12 July 1991. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S147466701750628X

[102] T. Morimoto, T. Takeuchi, and Y. Hashimoto, "Growth optimization of plant by means of the hybrid system of genetic algorithm and neural network," in *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, vol. 3. IEEE, 1993, pp. 2979–2982.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 249-255 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

A Smart Box for Blood Bags Transport: Simulation Model of the Cooling Autonomy Control System

Lorenzo Damiani*, Roberto Revetria, Emanuele Morra, Pietro Giribone

Genoa University, DIME (Department of Mechanical Engineering), 16145, Genoa, Italy

ARTICLE INFO

Article history:
Received: 14 November, 2019
Accepted: 08 January, 2020
Online: 27 January, 2020

Keywords:
Healthcare Safety
Simulation Model
Biologic Material Transport

ABSTRACT

This paper faces the topic of blood products transport by applying simulation, control and data exchange technologies in order to implement a safe and efficient integrated transportation system. As known, the bio-chemical properties of blood and its derivatives are subjected to deterioration whenever a well determined temperature range is not respected for an excessive time interval.

At present, blood products are mainly transported in thermally insulated boxes filled with ice which, other than ensuring a limited time of the material survival, are not able to allow any type of control over the state of the material and the history of temperature variations occurred within the storage volume.

The innovative proposal discussed in this paper regards a smart box for blood transport, equipped with a system for temperature control which uses elements of eutectic material stored in a separated compartment; eutectic exchanges heat with the transported blood bags through a separation wall on which blowers are installed for cold air circulation. The air flow is activated as soon as the box internal temperature increases over a pre-defined threshold, to extend as much as possible the eutectic material duration.

The smart box is equipped with sensors that dialogue by cloud with a central server on which is installed a Matlab-Simulink dynamic simulator of the box thermodynamic behavior, main object of the present paper. In particular, the simulator here presented has been improved by advanced heat exchange correlations and a more realistic system of thermal balance equations. The simulator allows to monitor the time remaining to the exhaustion of the cooling effect. In case of lack of autonomy, a logistic algorithm provides the coordinates of the nearest healthcare location for safe storage.

1. Introduction

This paper is an extension of the work originally presented in 2019 Spring Simulation Conference (SpringSim), Tucson, Arizona, USA [1]. The paper presented the study of a safe and integrated transport system for blood and blood derivatives, employing the most advanced simulation and communication technologies. In the present extended version, the simulator of the box dynamic behavior has been improved by implementing a more detailed and more realistic thermal exchange model.

The paper addresses the topic of healthcare safety [2-4], focusing on the problem of blood bags transport. The transportation of biologic material [5,6] presents risks related to the exceeding of deterioration temperature thresholds therefore, other than designing a transport container provided with a long range cooling system, it is necessary to manage the temperature control in different ambient conditions and to monitor the material status during travel time [7]; this aspect highlights the importance of data exchange with a central unit (hospital or blood bank), which is in charge of providing real-time logistic coordinates (i.e. the nearest possible point of blood collection) to prevent deterioration.

The usual procedures applied to the transport of biologic material provide the samples collection from distributed locations and their subsequent delivery to a laboratory for the analyses. In recent years, the increased costs and the complexity of clinical analysis equipment, has directed the management choices towards the concentration of the most important activities (e.g. analysis) in large centralized laboratories, from which depend a great number of smaller collection points. Only large and highly automated laboratories can reach the required scale economies to process a large number of samples with satisfying economical profitability.

*Lorenzo Damiani, Genoa, Italy. Email: Lorenzo.Damiani@unige.it

Since the distance between central labs and collection points can be significant and the traffic situation can dramatically increase transport time [8-10], the need to extend and control the autonomy of the cold storage becomes a central issue. Moreover, another critical aspect regards the continuous increase of blood banks centralization in order to cover larger areas, which involves longer travel time; thus, the issue arises of ensuring protection, traceability and continuous monitoring of blood bags during their travel.

The transport box ensures the cooling effect to the biologic material by exploiting an array of eutectic elements opportunely disposed in a separate compartment. From the latter, on-off blowers, activated by measured temperature signal, deliver a flow rate of cold air to the blood bags compartment, withdrawing heat to keep temperature within the safety range. The box envisages a cloud connection with a central server unit, on which is uploaded the dynamic simulator main object of this article, able to calculate the box temperature and thermal autonomy in function of time. On the same central server, a logistic model receives in real time the box geo-localization data [11-13] and gives as output the coordinates of the nearest center for safe blood storage in case of box thermal autonomy exhaustion.

The objectives of this paper are:

1) Implementing a more realistic thermal exchange model for the container, including air-blood and air-eutectic thermal exchange correlations;

2) Evaluating the dynamic characteristics of the container, by building diagrams which show the time variation of temperature for air, blood and eutectic and liquid mass fraction variation with time for eutectic;

3) Assessing the simulated performance of the container in terms of duration of the cooling effect and number of blowers interventions in different environmental conditions.

2. Features of the Blood Transport System

The main scope of the proposed integrated transport system is enabling healthcare operators to monitor, track and certify the quality of each transported biological good (plasma, blood, platelets…), by a complete traceability of the transport.

The integrated transport system consists in the active interaction between thermally insulated intelligent transport containers for blood products storage and a central computer unit dedicated to recording the whole history of the transport variables, other than processing the data received from the smart boxes (e.g. internal temperature and spatial position) to predict the exhaustion time and the nearest point of collection where the material can be safely stored. The transport boxes in motion dialogue with the central unit by cloud connection.

2.1. Smart Box Constructional Features

The smart container is constantly cloud-connected to a central server in order to monitor, record and signal, in real time, possible failures and alerts for the main operational variables, as:

- Box inner temperature;
- Ambient temperature;
- Relative humidity;
- Geo-localization (GPS, GSM, GLONASS, GALILEO);
- Speed of traveling;
- Accelerations;
- Tilting of the box;
- Physical integrity of the device;
- Date, time, minute
- Name of the responsible for each operation (switching on, loading, shipping, delivering, unloading, switching off, inspection).

In particular, regarding the latter function, each operator will be equipped with a card by means of which he can interact with the box; according to the field of expertise of the operator, he will be allowed to all or just certain actions on the box (e.g., a paramedic in charge for the transport will not be allowed to open the box or change the temperature set-point, whereas a medicine doctor will).

The box is designed to keep the inner temperature within a safety range, defined by the user according to the type of product and the external conditions. The cooling effect must contrast the thermal flow which enters the box from the environment and is ensured by eutectic plates confined in a separate compartment from the blood storage area. The eutectic plates release their thermal content when required, thanks to the “smart wall” which separates the eutectic plates zone from the blood products storage zone, providing the ventilation through correctly sized blowers activated by the signals from the sensors mounted inside the box. A scheme of the box components is shown in Figure 1.

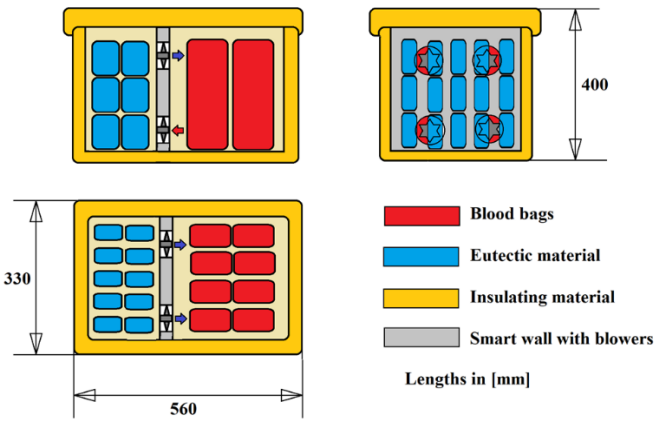


Figure 1: Schematic view of the Smart Box.

The box cooling system is not intended to cool down the biologic material starting from ambient temperature; therefore, the transported blood products need to be loaded at the desired transport temperature. The box is designed to keep the temperature range for no more than 5 hours.

The box functions are settable through a PLC equipped with LCD monitor. Operators interacting with the smart box are identified by a magnetic card which, basing on the permissions released to each person, permits to enable or disable the various

box functions (opening, closing, change of settings …). The box can also be managed through an online portal from which the user can monitor the status and interact with the boxes in travel.

2.2. Data Exchange

The above listed physical quantities are measured by means of sensors installed in the Smart Box, sampled with a frequency of 5 Hz. Such quantities are sent to a central server. According to the availability of GSM network, the collected data are stored in a local memory from which they are sent to the server. If available, the box can exploit a dedicated Wi-Fi router in place of the GSM.

In Figure 2 is shown the scheme of the network connections for data exchange.

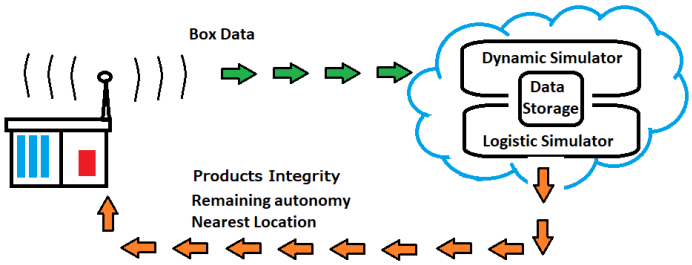


Figure 2: Box connection system.

3. Matlab Simulink Model

The dynamic modelling of a process has the purpose to reproduce, using differential equations, the variation with time of the system main parameters [14-17]. The variables observed by the smart box blood transport model are the blood temperature and the eutectic liquid mass fraction, which are influential parameters for the blood status and for the box autonomy.

Compared to the simulation model described in [1], the system of differential equations solved by the simulator here presented has been improved by introducing the thermal exchange equations between the air contained in the box and the stored blood and eutectic material. The algorithm for switching on-off the blowers has been set to have as controlling variable the temperature of the air contained in the box, measured by thermocouples.

3.1. Air Thermal Balance

Equation 1 describes the thermal balance of the air contained within the smart box, in contact with the blood, the eutectic, the confinement walls and the electric blowers for its circulation. The temperature of the air mass varies in function of the thermal flow contributions to which it is subjected.

$$\dot{q}_{tot} + \dot{q}_{eu} + \dot{q}_{bl} = M_{air} C_{p\_air} \frac{dT_{air}(t)}{dt} \qquad (1)$$

In Equation 1:

$T_{air}(t)$ is the temperature of the air contained within the box, of mass M_{air}, having specific heat capacity C_{p_air}, the latter calculated in function of temperature by using the Langen coefficients.

$\dot{q}_{eu}$ and $\dot{q}_{bl}$ represent the thermal flows exchanged between the air and, respectively, the blocks of eutectic material and the blood bags.

$\dot{q}_{tot}$ represents the total external flow heating up the material in the box, which is the sum of the thermal flow entering from the insulated walls ($\dot{q}_{ext}$) and the electric power of the blowers ($\dot{q}_{fan}$), as shown in Equation 2.

$$\dot{q}_{tot} = \dot{q}_{ext} + \dot{q}_{fan} = \\ = K_{box} S_{box} (T_{amb} - T_{air}(t)) \\ + \dot{q}_{fan} \qquad (2)$$

K_{box} and S_{box} are respectively the transmittance and external surface area of the box walls, T_{amb} is the external ambient temperature.

$\dot{q}_{fan}$ is the power of the four blowers (fans) installed in the smart wall to ensure the air circulation when required; since the air loop is closed, all the fans electric power (4 W) is injected into the air stream and contributes to increase its temperature when fans are active.

In Figure 3 is represented the scheme of the thermal balance for the air, implemented with Simulink blocks.

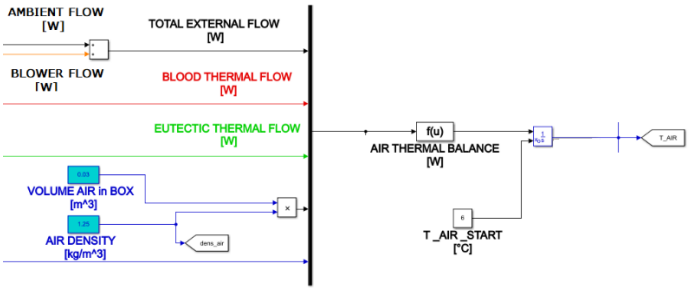


Figure 3: Simulink block scheme for thermal balance of air contained in the box.

3.2. Eutectic Thermal Balance

Equation 3 and Equation 4 reproduce the thermal balance of the eutectic material; the first equation is employed in normal operation, when the eutectic has not yet molten down completely and represents the eutectic behavior during meltdown, describing the liquid mass fraction time variation ($X_{eu}(t)$). Such parameter increases from 0 to 1 while temperature remains at the equilibrium value of – 2°C; the second equation considers the temperature increase of the molten eutectic ($T_{eu}(t)$) subjected to the incoming thermal flow.

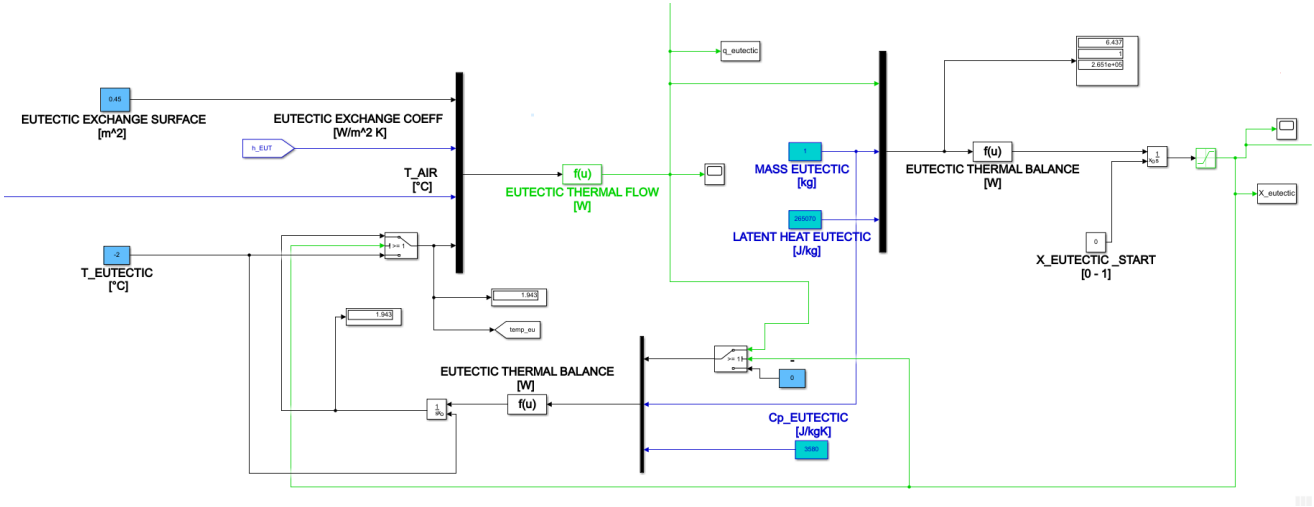
$$\dot{q}_{eu\_OP} = h_{eu} S_{eu} (T_{eu}(0) - T_{air}(t)) = \\ = M_{eu} r_{eu} \frac{dX_{eu}(t)}{dt} \qquad (3)$$

$$\dot{q}_{eu\_EX} = h_{eu} S_{eu} (T_{eu}(t) - T_{air}(t)) = \\ = M_{eu} C_{p\_eu} \frac{dT_{eu}(t)}{dt} \qquad (4)$$

In these equations:

M_{eu} is the eutectic mass, set to 1 kg;

r_{eu} is the latent meltdown heat of the eutectic, set to 265 070 J/kg;



C_{p_eu} is the specific heat capacity of the liquid eutectic, set to 3560 J/kg K;

h_{eu} is the thermal exchange coefficient between air and eutectic material blocks, the latter having a total heat exchange surface equal to S_{eu}.

$X_{eu}(t)$ is the liquid mass fraction of the eutectic, defined in Equation 5:

$$X_{eu}(t) = \frac{M_{liq}}{M_{liq} + M_{sol}} \tag{5}$$

M_{liq} and M_{sol} are the masses of liquid and solid eutectic during transport time.

Figure 4 shows the block scheme implemented in Simulink for the eutectic balance.

3.3. Blood Thermal Balance

Equation 6 represents the thermal balance of the blood contained in the storage vane.

$$\dot{q}_{bl} = h_{bl} S_{bl} (T_{air}(t) - T_{bl}(t)) = M_{bl} C_{p\_bl} \frac{dT_{bl}(t)}{dt} \tag{6}$$

In Equation 6:

$T_{bl}(t)$ is the blood temperature; the correct range for transport is between 4°C and 8°C.

M_{bl} is the mass of the stored blood, equal to 7 kg;

C_{p_bl} is the blood specific heat capacity, equal to 3900 J/kg K;

h_{bl} is the thermal exchange coefficient between air and blood bags, the latter having a total heat exchange surface equal to S_{bl}.

Figure 5 shows the block scheme implemented in Simulink for the blood balance.

3.4. Heat Exchange Coefficients

To compute the heat exchange dynamics between the air contained in the box and the blood and eutectic materials, it was necessary to carry out a research regarding the thermal exchange

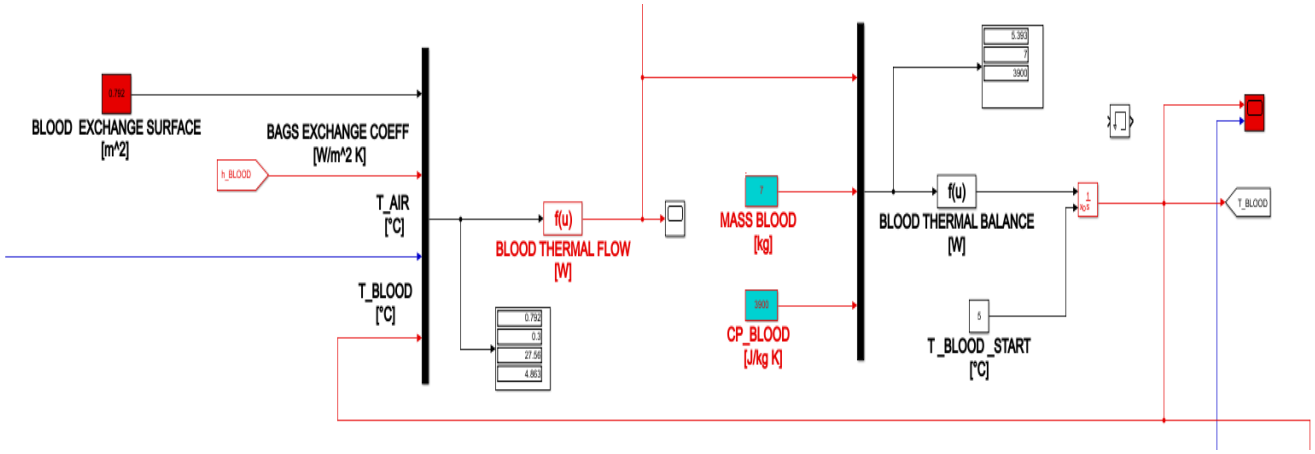


Figure 5: Simulink block scheme for thermal balance for thermal balance of blood material.

coefficients in confined narrow channels. [18-21]. Since the heat exchange within the box may occur both in presence of forced air circulation (blowers on) and in calm air (blowers off), two heat exchange correlations have been individuated and applied to the box geometric and thermo-fluid dynamical characteristics. The correlations employed, based on Nusselt number, are presented respectively in Equation 7 for the forced circulation case and in Equation 8 for the natural circulation case:

$$Nu = 0.036\, Re^{0.76} Pr^{1/3} \tag{7}$$

$$Nu = \left(\frac{144}{Ra^2} + \frac{2.873}{Ra^{0.5}}\right)^{-0.5} \tag{8}$$

Re being the circulating flow Reynolds number, calculated basing on the mass flow rate of the blowers, equal to 0.018 kg/s; Pr being the air Prandtl number and Ra the Rayleigh number, for natural convection correlations.

These two correlations have been implemented in Simulink blocks both for the blood compartment and for the eutectic compartment, determining the heat exchange coefficients in natural and forced convection; the switching between the two conditions occurs in function of the on-off signal provided to the blowers so that, when the latter are switched on, the heat exchange coefficient used by thermal balance equations is the one obtained by Equation 7; otherwise, the one obtained by Equation 8.

3.5. Temperature Control Modelling

Figure 6 illustrates the scheme, implemented in Matlab-Simulink, of the box temperature control system [22], finalized to keep the blood compartment in the safe temperature range.

The controlled variable is the box internal air temperature, which is detected by sensors and sent to the control system.

The controlling variable is the on-off signal to the blowers, managed according to the following logics:

Blowers switch-off as $T_{air}(t) < Tmin$;

Blowers switch-on as $T_{air}(t) > Tmax$;

Keep blowers status (on or off) as temperature remains in the desired range: $Tmin < Tair(t) < Tmax$.

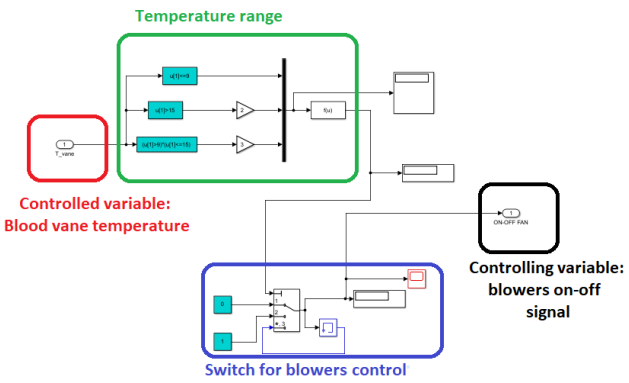


Figure 6: Matlab-Simulink block scheme of the blowers control system.

The control algorithm has been implemented by Matlab-Simulink logic blocks: the time varying compartment temperature is compared to the set-point values Tmin and Tmax, providing 0 or 1 signals which are sent to a multi-port switch that provides the on-off signal to the blowers respecting the above-mentioned logics.

4. Results of the Simulations

The smart box simulator improved with air-material thermal exchange correlations has been tested to assess its behavior. The main difference between the present model and the one proposed in [1] stands in the explicit modeling of the air behavior inside the compartments, operating as a heat vector during transport time.

The controlling variable selected to provide a feedback to the blowers on-off signal is the internal air temperature, while in [1] the blood temperature had been chosen. The control algorithm has been set in order to switch on the circulation blowers as air temperature approaches 8°C and to switch them off as it reaches 4°C. It has to be noticed that the blood loses its properties as its temperature overcomes 8°C for a significant time period.

Regarding the environmental conditions of the test, two different periods of the year have been chosen, namely May and August, characterized by external air temperatures of respectively 21°C and 27°C.

The model outputs monitored during simulations are:

- The blood temperature, $T_{bl}(t)$;
- The temperature of air contained in the box, $T_{air}(t)$;
- The liquid mass fraction of the eutectic $X_{eu}(t)$, in the [0 – 1] range;
- The eutectic temperature $T_{eu}(t)$; as soon as $X_{eu}(t)$ reaches 1, the eutectic latent heat useful for cooling is exhausted; however, some cooling is still ensured by the eutectic sensible heat, while increasing its temperature;
- The number of blowers on-off cycles;
- The time to cooling effect exhaustion.

The following paragraphs illustrate the simulations results; the tests have been performed setting the input parameters listed in the 3rd chapter and the simulator running time has been stopped after about 7 hours (25 000 s).

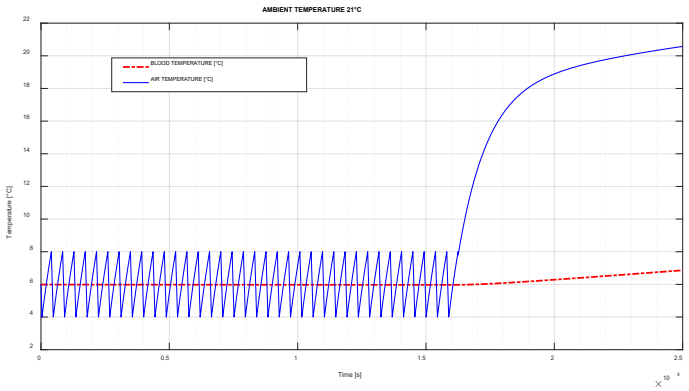


Figure 7: Blood and air temperatures in function of time (simulation for the month of May).

4.1. Results for the Month of May

Figure 7 shows the diagram of blood and air temperatures in function of time.

The blood remains in the correct temperature range for all the simulated time period, reaching a temperature of 6.9°C at the instant of simulation stop (25 000 s, about 7 hours). The circulation blowers are switched on for 36 times, as verifiable from the air temperature curve. As soon as the eutectic reaches exhaustion, internal air temperature increases tending to the ambient temperature value.

Figure 8 shows the time diagram for eutectic temperature and liquid mass fraction.

The eutectic effect ends at 16250 s (4.51 h), after which its temperature increases from the equilibrium value of – 2°C and tends, after a transient dictated by the material thermal capacity, to reach equilibrium with the external temperature.

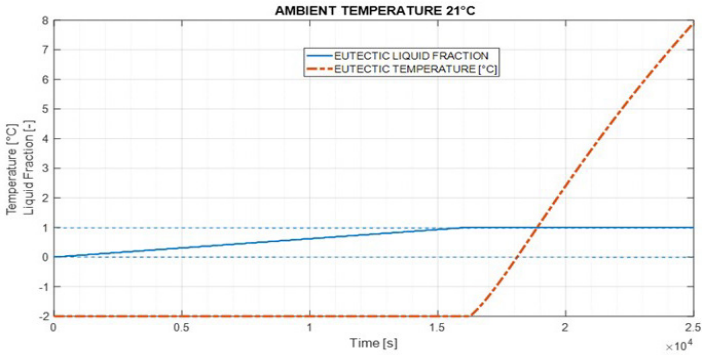


Figure 8: Eutectic liquid mass fraction and temperature in function of time (simulation for the month of May).

With respect to the results obtained in [1], the increase in the number of blowers on-off interventions appears remarkable. This effect is due to the inclusion of the air-materials thermal exchange effect: the mass of air contained in the box is limited (0.03 kg) and its specific heat is low compared to liquid or solid substances; therefore, air temperature rises and falls result very fast with respect to blood, which keeps almost constant temperature for all the simulated period. Having chosen the internal air temperature as control variable for the blower signal, the latter is forced to follow the air behavior, which increases dramatically the number of on-off switches.

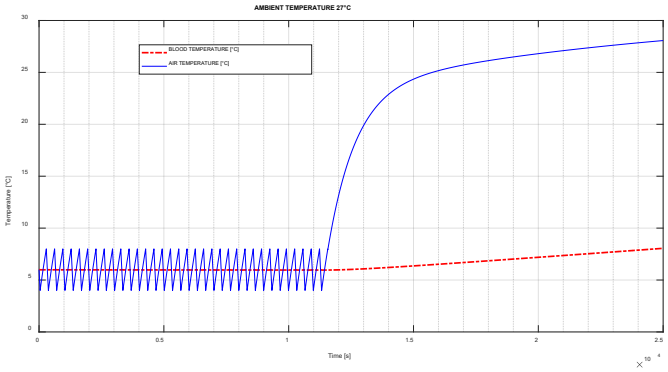


Figure 9: Blood and air temperatures in function of time (simulation for the month of August).

4.2. Results for the Month of August

Figure 9 reports the diagram of blood and air temperatures in function of time.

In the hot month of August, the blood temperature keeps below 8°C (blood deterioration threshold) for 24 701 s (6.86 h), and the blowers are activated for 34 times. As soon as the eutectic reaches exhaustion, internal air temperature increases tending to 27°C.

In figure 10 is shown the eutectic temperature and liquid mass fraction diagram in function of time. The eutectic cooling effect is exhausted after 11330 s (3.15 h).

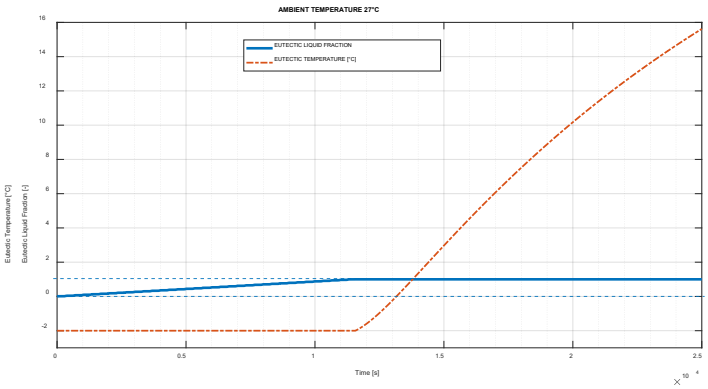


Figure 10: Eutectic liquid mass fraction and temperature in function of time (simulation for the month of August).

Table 1 reports a synthesis of the simulations parameters and results for the two periods considered.

Table 1: Synoptic table of the simulated parameters.

Parameter	[Units]	May	August
Ambient temperature	[°C]	21	27
Stored blood mass	[kg]	7	7
Initial blood temperature	[°C]	5	5
Eutectic mass	[kg]	1	1
Eutectic equilibrium temperature	[°C]	-2	-2
Blowers flow rate	[kg/s]	0.018	0.018
Eutectic effect exhaustion time	*[h]*	*4.51*	*3.15*
Threshold temperature attainment time	*[h]*	*> 7*	*6.86*
Blowers interventions	*#*	*36*	*34*

5. Conclusions

Blood products transport, topic of clear interest for the healthcare sector, has been addressed in the original article [1], which proposed a thermally insulated portable box equipped with a ventilation cooling system and an informatics support system monitoring the transport data (e.g. time remaining before the exhaustion of cooling effect, the logistic data of the nearest blood collection point…) through a central control unit. The present extended paper addresses the improvement of the dynamic Matlab-Simulink simulator for said blood transportation system.

The Matlab-Simulink model is supposed to help the correct sizing of the box in a design phase (walls insulation, mass ratio

between eutectic and blood, blowers air flow rate, …) and to operate as an online simulator for calculating the box main variables (blood temperature and remaining autonomy) during transport.

The improvement on the Matlab-Simulink model, which regarded the introduction of correlations for considering the thermal exchanges between air and transported materials, is expected to provide results more in accordance with real operation.

The test simulations have been carried out for two periods of the year, namely May (21°C) and August (27°C). In general, the results of the improved model showed a similar cooling effect exhaustion time compared to the original model in [1]. The main difference stands in the number of interventions of the controlled blowers, which resulted about 7 times more frequent than what predicted by the original model. This effect results from the inclusion of the internal air heat exchange correlations and from the choice of using air temperature as the controlling variable for the blowers activation; being the small amount of air trapped within the box very fast in increasing and decreasing temperature, it results clear that the blowers will be called in operation frequently.

Future development of the technology described in this paper is the construction and testing of a smart-box for blood products transport, in order to assess its real performance when interacting with the actual environment. Also, more effort will be addressed to the development of the inter-connection informatics system, together with the implementation of an efficient geo-localization and optimization algorithm for the indication of the nearby healthcare facilities.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Authors would like to acknowledge Genoa University.

References

[1] Damiani L., Revetria R., Arecco S., Simulation model of the control system of portable boxes for blood bags transport, SpringSim-MSM 2019, April 29-May 2, Tucson, AZ, USA. DOI: 10.22360/springsim.2019.msm.001.

[2] Damiani, L., Revetria R., Mzoughi K., Giribone P., A Hybrid Simulation Model for Evaluating Risk Related to Complex Plants in Hospitals, Proceedings of the World Congress on Engineering and Computer Science 2016 Vol II WCECS 2016, October 19-21, 2016, San Francisco, USA.

[3] Damiani, L., Giribone P., Mzoughi K., Revetria R., A Hybrid Simulation Model for Hospital Complex Plants Risk Evaluation, 2017, Engineering Letters, 25.2.

[4] Patrone, C., Cassettari L., Damiani L., Mosca R., Revetria R., Optimization of Lean Surgical Route through POCT Acquisition, IAENG Conference IMECS 2017, Hong Kong, 15 – 17 March 2017.

[5] Ruiz-García, L., et al.: Monitoring the intermodal, refrigerated transport of fruit using sensor networks. Spanish Journal of Agricultural Researh 5(2), 142–156 (2007).

[6] Gil-Castiñeira, F., Chaves-Diézquez, D., González Castaño, F.J.: Integration of nomadic devices with automotive user interfaces. IEEE Transactions on Consumer Electronics 55(1), 34–41 (2009).

[7] Castro, M., A. J. Jara, A. F. G. Skarmeta, Analysis of the Future Internet of Things Capabilities for Continuous Temperature Monitoring of Blood Bags in Terrestrial Logistic Systems, in: Lee G., Howard D., Ślęzak D. (eds) Convergence and Hybrid Information Technology. ICHIT 2011. Lecture Notes in Computer Science, vol. 6935, Springer, Berlin, Heidelberg.

[8] Gibson, J.G., T. Sack, R.D. Evans, W.C. Peacock, The Effect of Varying Temperatures on the Post Transfusion Survival of Whole Blood During Depot Storage and after Transportation by Land and Air, The Journal of clinical investigation, 1947, Vol.26(4), pp.747-755.

[9] Or, I., W.P. Pierskalla, A Transportation Location-Allocation Model for Regional Blood Banking, A I I E Transactions, 11 (2), 1979.

[10] Fakhr, A.A., Bikmoradi A., Zandiyh M., Mozafari H., The assessment standards related to blood transportation, storage, preparation, and administration in operating rooms of Hamedan educational treatment centers, Scientific Journal of Iranian Blood Transfusion Organization, 2012, Vol. 9 Issue 1, pp. 63-71.

[11] Ngai, E.W.T., Cheng, T.C.E., Au, S., Lay, K.: Mobile commerce integrated with RFID technology in a container depot. Decision Support Systems 55(1), 62–76 (2007).

[12] Atzori, L., Iera, A., Morabito, G.: The Internet of Things: A survey. Computer Networks 54, 2787–2805 (2010).

[13] Guinard, D., et al.: Interacting with the SOA-Based Internet of Things: Discovery, Query, Selection, and On-Demand Provisioning of Web Services. IEEE Transactions on Services Computing 3(3), 223–235 (2010).

[14] Damiani, L., Trucco, A. An experimental data based correction method of biomass gasification equilibrium modeling, Journal of Solar Energy Engineering, 2010, 132(3).

[15] Damiani, L., Giribone P., Revetria R., Pini Prato A., A passive decay heat removal system for the lead cooled fast reactor demonstrator "Alfred", Progress in Nuclear Energy 83 (2015) 294-304.

[16] Damiani, L., Pini Prato A., Simulation of a Power Regulation System for Steam Power Plants, Energy Procedia, 45 (2014) pp. 1185-1194.

[17] Dellachà, J., Damiani L., Repetto M., Pini Prato A., Dynamic Model for the Energetic Optimization of Turbocompound Hybrid Powertrains, Energy Procedia 45 (2014) 1047 – 1056.

[18] Kotcioglu I., Ayhan T., Olgun H., Ayhan B., Heat Transfer and Flow Structure in a Rectangular Channel With Wing-Type Vortex Generator, Tr. J. of Engineering and Environmental Science 22 (1998) , pp. 185 - 195.

[19] Rybinski W., Mikelewicz J., Analytical solutions of heat transfer for laminar flow in rectangular channels, Archives of Thermodynamics, Vol. 35(2014), No. 4, pp. 29–42 DOI: 10.2478/aoter-2014-0031.

[20] Forrest E.C., Hu L.W., Buongiorno J., McKrell T.J., 2016, Convective Heat Transfer in a High Aspect Ratio Mini-Channel Heated on One Side, Journal of Heat Transfer, 138(2): 021704.

[21] Tanda G., Natural convection heat transfer in vertical channels with and without transverse square ribs, 1997, Int. J. Heat Mass Transfer, Vol. 40, No. 9, pp. 2173-2185.

[22] Damiani, L., Giribone P., Guizzi G., Revetria R., Romano E., Different approaches for studying interruptible industrial processes: Application of two different simulation techniques, Handbook of Research on Computational Simulation and Modeling in Engineering, 2015.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 256-262 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Hospital Warehouse Management during the construction of a new building through Lean Techniques

Carlotta Patrone*,[1], Maryna Mezzano Kozlova[2], Monica Brenta[3], Francesca Filauro[4], Donatella Campanella[4], Antonietta Ribatti[5], Elisabetta Scuderi[5], Tiziana Marini[6], Gabriele Galli[2], Roberto Revetria[2]

[1]*E.O. Ospedali Galliera, General Directorate, Mura delle Cappuccine 14, 16128, Genoa, Italy*

[2]*University of Genoa, Mechanical, Energetic, Industrial and Transport Engineer Department (D.I.M.E.), Via Montallegro 1,16145 Genoa, Italy*

[3]*E.O. Ospedali Galliera, Anaesthesia Department, Mura delle Cappuccine 14, 16128, Genoa, Italy*

[4] *E.O. Ospedali Galliera, Pharmacy, Mura delle Cappuccine 14, 16128, Genoa, Italy*

[5] *E.O. Ospedali Galliera, Warehouse and Supply Department, Mura delle Cappuccine 14, 16128, Genoa, Italy*

[6] *E.O. Ospedali Galliera, Information and Technology Department, Mura delle Cappuccine 14, 16128, Genoa, Italy*

ARTICLE INFO

Article history:
Received: 22 October, 2019
Accepted: 23 December, 2019
Online: 27 January, 2020

Keywords:
Business Process Reengineering (BPR)
Lean
Management
Warehouse
Supply Chain Resiliency
Design of Experiments
Healthcare

ABSTRACT

This paper shows a method applied during the reconstruction of an existing middle size Italian hospital. In fact, during this delicate phase, the project had a significant decrease in storage space and access gates for goods, medicines and medical devices. This aspect has a negative impact on road traffic. There was no consensus among the management on how to deal with this problem. In addition, a great tension between engineers and warehouse managers made it impossible to find a solution. For the best of our knowledge there is no available literature on this topic. Hence, a deep analysis had been carried out according to the Business Process Reengineering (BPR). For this reason, a hospital lean group has been created and a statistically study on storage volume inside the warehouse and a monitoring of trucks and goods arrival has been implemented. Three different solutions had been identified through data collection, design of experiments and lean application. The first one was focused on an exclusively internal management, the second one on an exclusively external management whereas the last one was focused on a partially internal scenario. Advantages, disadvantages, costs and the impact on the road traffic has been highlighted for each scenario. Afterward the Management had all the elements to make the best decision. The method reported in this study could be extended during the construction or reconstruction of any company that needs a warehouse or to face a problem in a tense atmosphere.

1. Introduction

Warehouse management inside hospitals has always been considered a crucial topic. Venkateswaran highlighted the complexity of the management of drugs warehouses for the uncertain demand and the high risk of product unavailability that could potentially compromise the safety of the patients [1]. One of the more critical issues is the difficulty to apply Lean Technologies inside the warehouse. According to Frank, warehouses and Lean are considered as mutually exclusive due to the strong difficulties of application of this method in a stockage setting. Nonetheless, the potential advantages are to reduce the "non-value-added steps" as much as possible and to optimize the stock flow [2].

This paper has been developed during the construction of a new building and the main target subjects were the office supply, the medical devices and the drugstore warehouse of an Italian highly specialized hospitals (430 beds). Moreover, during this period, the warehouse and the entry have been involved in the building site. More in detail, the project had a significant decrease in storage

*Corresponding Author Carlotta Patrone, E.O. Ospedali Galliera (Italy), patronecarlotta@gmail.com, +390105632975

space and access gates for goods, medicines and medical devices to the hospital had declined dramatically with an intense impact on the road traffic. In addition, there was no consensus among the management on how to deal with this problem. Furthermore, a great tension between engineers and warehouse managers made it impossible to find a solution. No study in a similar setting has been found in the literature. Therefore, it was necessary to analyze the most appropriate solution for the warehouse management during the construction phase to give to the management all the elements to choose the best solutions.

A deep analysis has been performed and then structured into two different phases, according to Business Process Reengineering (BPR). This paper is organized into five sections. Section 2 provides a literature review of the three main techniques used to develop this study. The methodology followed during the new building construction has been introduced in section 3. Section 4 illustrates the results obtained whereas the conclusions have been presented in section 5.

2. Literature Review

A review of the literature has been performed through the following keywords: “supply chain resiliency”, “lean healthcare” and “lean warehouses”.

2.1. Supply Chain Resiliency

Supply Chain Resiliency has been defined in a multidisciplinary view as “the adaptive capability of the supply chain to prepare for unexpected events, respond to disruptions, and recover from them by maintaining continuity of operations at the desired level of connectedness and control over structure and function” [3]. From a different point of view, another definition is “The ability of a system to return to its original state or move to a new, more desirable state after being disturbed” [4]. A resilient supply chain must respect the following principles:

- Resilience: it must be part of the initial project and the base of any decision. In other words, it has to be transversal to each side of the company. It is essential a high grade of risk collaboration and the resulting agility in the processes and their connection.

- Reactivity: The company must be ready to quickly respond to the unpredictable events in the most optimal way possible [5]. An efficient company has to work continuously on the vulnerability reduction, decreasing the probability of breakage and enhancing the resilience of their processes. This may occur by increasing the redundancy (increasing the cost) or the flexibility based on the organization of day-to-day operations [6]. Frequently the decision concerning the use of the resilience engineering method and Lean technique is doubtful [7].

2.2. Lean Warehouse Management in healthcare

The Lean has found widespread use in the healthcare sector. Its first hospital application occurred in 2001 in the United Kingdom and the next year in the United States [8]. In 2009, in Canada, five organizations found a way to increase the organization with outstanding results thanks to Lean. Consequently, it has been possible to decrease the waiting time both for the emergency rooms and for medical visits. Moreover, the operating and radiological room utilizations has been enhanced as well as the infectious diseases control [9].

Lean is an appropriate technique to tackle the materials inside the hospitals. A comparison study between the traditional storage and the Lean approach applications has been performed. The collecting data through questionnaires has been used and further statistical analysis has been performed [10]. Another suitable technique is the VIC strategy. This method is based on a mapping of the values flow (V), an integration of the stock information (I) and a final control (C) on these [11]. The flow mapping follows the production path from its start to its end and it shows a visual representation of the material flows and their information in each process [2].

In order to obtain better inventory management performances, the Customer Managed Inventory / Inventory Managed Inventory method (CMI/VMI) could be introduced inside the hospital. VMI/CMI are cooperation forms between a supplier and a client [12].

An improvement system of the warehouse management has been introduced in a hospital in Michigan and it is called “two-bin supply chain system”. For each specific medical supply, two containers were provided. This system has been designed in order to reduce the necessary time for the warehousemen to perform the inventory and to decrease the trends of an excessive order [13]. Still inside the hospital warehouse a comparison study between the 5S traditional and the 5S hybrid process has been implemented. 5S is a sequence of activities (sort, set in order, shine, standardize and sustain) and it is demonstrated as an effective tool for the supply management or organization. On the other hand, the hybrid 5S contains inventory management and process improvement tool techniques [1]. In order to manage the warehouses in a quicker and more accurate way it is possible to reinforce the Lean techniques using Radio Frequency Identification (RFID) technologies [14].

Studies focused on the warehouse personnel and on their sensitization about the downtimes have been treated. This waste of useful time could be used for other important procedures inside the hospital [10]. In order to sensitize the personnel, it is possible to introduce the Continuous Positive Reinforcements (CPR) technique. This is a management system using continuous positive reinforcement to achieve the objectives of the working groups [15].

The hospital warehouse management through the Lean application has been successfully described in study developed in province of Parma [16]. Two cases have been analyzed. The purpose of the first one was the cooperation among the hospitals in the region, including an exchange of drugs, medical devices etc., utilizing the same archive software in order to optimize the reorder and the storage process inside each hospital. The second case analyzed was an external centralized warehouse for all the hospitals in the province [16].

Another system adopted in China is based on a partial externalization of the hospital warehouses. It’s called the inventory

pooling. In other words, the hospital maintains the inventory minimum level, normally for 1 or 2 days of demand while most of the inventory is retained in the distribution center. The external warehouse organizes the daily deliveries of the drugs based on the information shared real-time by the hospital [17].

2.3. Business Process Reengineering (BPR)

The BPR has spread above all in the Anglo-Saxon world in the early 90s in the private firm sector. In those years, one of the first definition of process reengineering has been provided by Hammer in a paper appeared on Harvard Business Review [18]. In this definition, the author defines this technique as "the substantial afterthought and the radical redesign of the business processes for the purpose to obtain relevant results in terms of quality and services improvements to the user and enhancement in the productivity (cost reduction in view of a volume rise of the performance provided)" [19]. Later, around the middle of 90s, some publications based on this approach started also in the healthcare sector. The BPR in a managerial approach for the management of the organizational change based on the logic for processes.

3. Methodology

One of the key success factors of the BPR implementation is the human involvement [20]. For this reason, a dedicated hospital Lean group has been created. The group is composed of warehouse managers, industrial engineers, civil engineer, architect, nurses, physician, IT technician, and pharmacists.

The authors followed the BPR steps.

3.1. AS IS

The first step was to detect the actual warehouse management with the relative data. The Lean group mapped the flow through the Value Stream Mapping (VSM) technique. Thereafter, a data collection has been implemented in order to detect:

- actual storage space and the number of goods access available;
- real utilization of the dedicated space and of the access.

The actual storage space is 1.631, 25m2 with two dedicated goods access available. It seems to be an easy task but the actual storage spaces are scattered and they are managed by different heads.

For the second purpose a six weeks data detection on storage volume trend and truck arrival has been carried out.

For an accurate data collection, the warehouses have been divided in different areas. Moreover, a monitoring of the truck arrival and departure times, the volume and the timing of the merchandise arrival and the various necessary controls have been performed. Microsoft Excel tables have been created to achieve an efficient collection.

Figure 1 shows a focus on the volume storage trend and Figure 2 illustrates the comparison between standard deviation, mean and the true value of the volume.

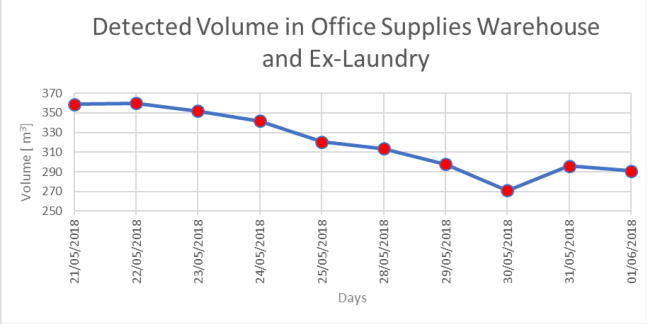


Figure 1: Storage in Office Supplies Warehouse and Ex-Laundry.

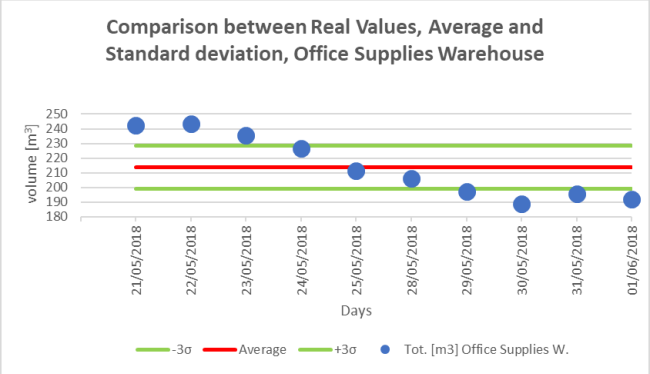


Figure 2: Comparison among Real Values, Average and Standard Deviation, office supplies warehouse.

An additional study based on the filling percentage of the space has been performed and reported in Table 1. This means that many area and shelves do not occupy most of the available space. In parallel to this phase, the template for the arrival of trucks has been filled in. Two different good access are available: one dedicated to the office supply warehouse and other to the pharmacy.

Table 1: Percentage of filling, current state of the warehouse

	[m^3] occupied	[m^3] available	% occupied
Shelves	80,09	273,15	29%
Corridor and Atrium	20,84	74,54	28%
Pallet Rooms 003-007-009	140,41	288,54	49%
Ex-Laundry Room	116,181	398,92	29%

A statistical analysis has been performed to understand if the time slot or the day of the week or the week affect the truck arrival. Latin Square is applied when disturb factors occur in addition to the treatments and to the first disturb factor [21].

Concerning the data analysis of the truck monitoring of the office supply warehouse, the results of the Latin Square Method highlighted a significant difference in the truck arrival by changing the time slots. Day and week did not affect the truck arrival. The results are reported in Table 2.

Table 2: Latin Square Method

Source of Variation	F_0	$F_{0,05;4;12}$	Conclusions
Weeks	0,81	3,26	$F_0 < F_{0,05;4;12}$
Time slots	7,70	3,26	**$F_0 > F_{0,05;4;12}$**
Days	2,27	3,26	$F_0 < F_{0,05;4;12}$

Analysis of Variance (ANOVA) has been performed on that data because only time slots affected the results. ANOVA is a set of statistic techniques whom are part of the inferential statistic. Moreover, they allow to analyze two or more groups of data by comparing the intern variability of these groups with the variability among these [22]. The ANOVA results for the office supply warehouse has been reported, as an example, in the Table 3 and 4. The two most important factors in using ANOVA to accept or reject the null hypothesis were calculated.

Table 3: Anova Results of Office Supply Warehouse (1/2)

Source of Variation	SS	Gf	MS
Time slots	137,62	10	13,762
Days	7,03	4	1,758
Interaction	34,10	40	0,853

There is a statistical significance to the results if F-value (F) is larger than the F critical value (F crit). Also, these results showed that only time slots have the statistical significance as shown in Table 2. The p-value is less than 0.05, so was rejected the null hypothesis that there's no difference between the means and the conclusion is that a significant difference does exist.

Therefore, the One-Way ANOVA test has been performed to understand which time slots are relevant and to provide a ranking of them. The Duncan test has been carried out to define the order of importance among the times using the One-Way ANOVA results. The output of this test, shown in Table 5, has defined which time slots have significant differences to each other.

Table 4: Anova Results of Office Supply Warehouse (2/2)

Source of Variation	F	P-value	F crit	Conclusions
Time slots	**14,92**	3,4E-21	**1,87**	**F>Fcrit p-value< 0,05**
Days	1,91	0,11	2,40	F<Fcrit p-value> 0,05
Interaction	0,92	0,60	1,44	F<Fcrit p-value> 0,05

Table 5: Duncan's Multiple Range Test

Time slots	Treatments (significant differences in bold)	Differences between means	Shortest significant range
8:30-10:00	**6:30-8:30**	8,50	3,19
	10:00-10:30	7,50	3,53
	11:00-12:00	5,33	3,49
	10:30-11:00	3,50	3,41
10:30-11:00	**6:30-8:30**	5,00	3,53
	10:00-10:30	4,00	3,49
	11:00-12:00	1,83	3,41
11:00-12:00	6:30-8:30	3,17	3,49
	10:00-10:30	2,17	3,41
10:00-10:30	6:30-8:30	1	3,41

Different results are achieved by the pharmacy access. Indeed, there are no remarkable differences among the time slots, the days or the weeks there. The truck arrival has a rectangular distribution since the analysis shows that the randomness with which the couriers deliver is very high. Hence, the truck arrival in the drugstore warehouse is a completely random variable with a rectangular distribution.

3.2. TO BE

The future state is the warehouse management with a storage space of 618,8 m^2 and with only one good access for all the hospital. Indeed, during this phase, this gateway will be the unique access to the hospital instead of the actual three (two dedicated to the goods and one two the services).

The authors individuated three possible solutions:

1. *Scenario 1*: internal management of the warehouses.
2. *Scenario 2*: assigning the warehouses management to an external firm of logistic.
3. *Scenario 3*: renting of the new warehouse and the procurement of the logistic of the deliveries to the Hospital.

For all the three solutions the following parameters, reported in section 4, have been studied:

- Economic Analysis: the balance between the emerging and the loss costs.
- Impact on Traffic
- Advantages: the point of strength of the scenario.
- Disadvantages: the point of weakness of the scenario.

Scenario 1

This scenario consists in the completely internal warehouse management. There are two main problems in this context. The first is the 62% less of storage space and the second the unique access to the hospital.

A data extraction on a one-year time frame has been performed in order to calculate the rate of rotation of the goods. The future available space turned out much smaller than the actual one but, thanks to the results performed with the rotation index, it is possible to notice that by increasing this one all the material currently available in the medical devices and office supply warehouse can be incorporated. As shown in the Figure 3, the rotation index of many products is very small or even zero, this means that in a one-year time slot the products are not completely used and their supplies are left. After a Pareto Analysis is possible to establish which products can be moved more often, in order to reduce their stock in the warehouse and to occupy less volume accordingly. It is also possible to determine which products could be put in the higher and less reachable shelves and which ones could be put in the reachable shelves.

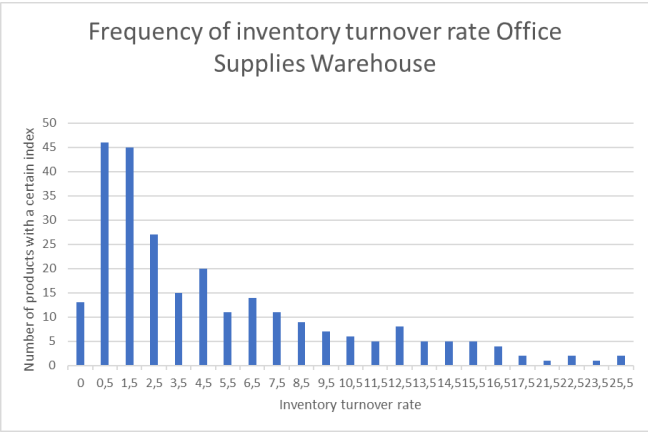


Figure 3: Inventory turnover rate, Office Supplies Warehouse.

Indeed, by increasing the filling percentage in the future shelves, it will be possible to optimize the spaces available and all goods, except drips (IV), can be allocated to warehouses despite the large reduction in space. The IV management will have to be outsourced.

Considering only the contemporary truck arrival of the office supply warehouse and drugstore it is possible to suppose a co-presence of about eight trucks in the rush hours. The future truck check-in ramp during the building of the new hospital will receive all the stocks instead of the actual organization. Furthermore, it will have to be the transit point toward many other stocks directed to other wards (e.g. radiopharmaceuticals, catering services and construction site). In order to understand more about the future traffic, an overlap of the trucks has been carried out and has been reported in Figure 4.

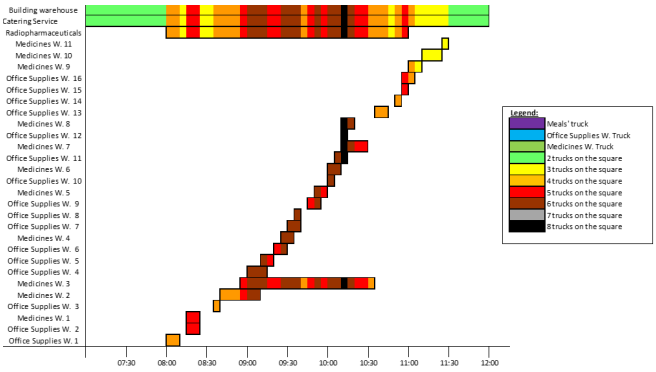


Figure 4: Occupation of the passage square, transitional state.

Figure 4 highlights the co-presence of at least two trucks. In addition, the street on this entrance is a one way one so this is a critical issue to consider.

All the costs involving in this scenario such as the costs for structural costs, outsourcing service for IV, etc. are reported in section 4.

Scenario 2

This scenario consists in the outsourcing of the service. This option comprises the delivery of the goods directly to the wards from the external warehouse. There would not be impact on traffic. Indeed, the goods delivery would happen once a day, at the time chosen by the hospital. The delivery time slot would be probably in the afternoon when the most part of the accesses to the hospital have already been done.

A market analysis has been carried out to understand the cost of this service. In addition, the impact on this scenario on the personnel actually dedicated to the warehouse has been studied with a reduction of them through:

- Pensions
- Termination of fixed-term contracts
- Transfer of personnel

Scenario 3

In this solution the hospital should rent a warehouse and the service of delivery of material from the rented warehouse to the Hospital. A market analysis has been carried out to understand the

cost of this service. The warehouse would be managed the hospital with a staff increase.

4. Results

The detailed results of the cost analysis implemented are reported in Table 6.

Table 6: Scenario Costs Analysis

Cost	Scenario 1	Scenario 2	Scenario 3
Staff	2.700.000€ (5 years x 30.000€/year x 18 workers)	750.000€ (5 years x 30.000€/year x + 5 workers)	2.700.000€ (5 years x 30.000€/year x 18 workers)
Additional or diminutive Staff	7500.000€ (5 years x 30.000€/year x 5 additional workers)	-1.9500.000€ (13 workers less*5years*30.000€/year)	750.000€ (5 years x 30.000€/year x 5 additional workers)
Construction costs	€ 823.858,00	€ 105.000,00	€ 275.150,65
Automated warehouses for drugs			
Wardrobes and shelving warehouse	€ 20.292,65		
I.V. outsourcing	500.000€ (5 years x 100.000€/year)		
Optical pen	€64,66		€64,66
Outsourcing fee		5.000.000€ (5 years x 1.000.000 €/year)	
Rental warehouse			318.000€ (5.300€/months per 60 months)
Rental delivery service from rented warehouse to hospital			2.122.800€ (348.000 €/years*1,22 VAT*5 years)
Total costs (5 years)	**€ 4.794.215,31**	**€ 3.905.000,00**	**€ 6.166.015,31**

Table 7 highlights the advantages and disadvantages of each scenario.

Table 7: Scenario Analysis

SCENARIO	ADVANTAGES	DISADVANTAGES
SCENARIO 1: The exclusively internal management	The directly availability of the stocks. Tendering procedure is not requested.	The rooms' utilization because they are very small. The manual handling is not optimal, the workforce is expected to increase. The medicines' reception and control will be problematic because of the questions that arise about the responsibility and the increase of staff.
SCENARIO 2: The exclusively external management	The solution is immediate	Tendering procedures and the consequently long time
SCENARIO 3: the partial use of internal structures and the rent of the external warehouse	The direct control over the material. There is no immediate expense for the renovation of the space necessary for the transitional warehouse.	Tendering procedures. The increase of the complexity in the management of remote staff and the probable increase of employees.

5. Conclusions

The completely intern solution (Scenario 1) has a very strong impact on the viability. One advantage of this solution is that the items are immediately available and there are no conflicts in the item management, shorter availability time and no tender procedure. The uniquely external solution (Scenario 2) has a low impact on the viability and it is immediate. There is a loss of control in the stocks, confusion in the role of the stock responsibilities etc.

The partially intern solution (Scenario 3) with the rent of an external warehouse has a low impact on the viability. The advantages are the direct control on the merchandise, lack of an immediate disbursement for the restructuring of the spaces. On the other hand, the drawbacks are the increasing in complexity of the remote employee management and a probable increase in personnel.

All the analysis has been subjected to the attention of the Direction. They decided to adopt Scenario 2.

This paper reports a method and it could be applied in all the contexts during the building of a new construction. In the hospital this approach has been used because the management didn't know how to manage the stocks during the building of the new hospital

because the spaces for the warehouse were limited and there was an impact on the traffic.

The authors implemented the BPR approach involving all the professionals. This approach achieved the goal because it gives all the elements to the Management to make a decision and all the people feel involved in this phase. In fact, at the end of the project the tensions initially present were completely eliminated.

One possible limitation of the study could be that the market analyses carried out on the August of 2018 could be may change depending on new solutions provided by companies. However, the management of the hospital needed to take a decision on that moment. Nevertheless, the aim of this paper is to provide a method and not an economic estimate. Future research on this topic could be done using discrete event simulation to have more detail on the impact of each solution on the system.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

A special thanks to Maria Laura Zizzo, Cinzia Gina Cravisana, Ezio Nicolas Bruno Urbina. Federica Solari, Robrta Clavarino, Mauro Nelli and Carla Elda Fraguglia for their important contribute to this paper.

References

[1] S. Venkateswaran, "Thesis. Implementing lean in healthcare warehouse operations - evaluation of 5S best practice" Louisiana State University and Agricultural and Mechanical College Master's Theses, 2011.
[2] P. Frank, C. Garcia, "Applying Lean Concepts in a Warehouse Operation" IIE Annual Conference and Exhibition. 2004. p. 2819-2859.
[3] S. Holcomb, "Understanding the concept of supply chain resilience" The international journal of logistics management, **20**(1), 124-143, 2009.
[4] Collins, H. (2000). Collins English Dictionary. Glasgow.
[5] H. Peck, M.C. Peck, "Building the resilient supply chain" The international journal of logistics management, **15**(2), 1-14, 2004.
[6] Y. Sheffi, J.B. Rice Jr. "A Supply Chain View of the Resilient Enterprise" MIT sloan management review, **47**(1), 41-48, 2005.
[7] T. A. Saurin, C. B. Rosso and A.L. Colligan "Towards a resilient and lean health care," In: Resilient health care. ROUTLEDGE in association with GSE Research, 2017. p. 3-17.
[8] H. Yalin, and T.E. Yonetimi, "Lean Medical Inventory Management in Hospitals," International Journal of Academic Value Studies (Javstudies), **3**, 16-24, 2017.
[9] B.A. Fine, B. Golden, R. Hannam and D. Morra, "Leading Lean: A Canadian Healthcare Leader's Guide," Healthcare Quarterly, **12**(3), 32-41, 2009.
[10] A. Yildirim, "Innovative approach to health service materials management" Hestourex World Health Sports Tourism Congress & Exhibition, 6-9 April 2017, Antalya (33-42).
[11] E.F. Ramly, "Lean Supply Chain contribution toward Lean Health Care," International Conference on Continuous Value Adding based on Lean Thinking in Hospital Management, Tehran, Iran, 30 Nov – 1 Dec 2011.
[12] L. Michailidou, "Healthcare supply chains: a case study of hospital vendor collaborative practices" International Journal of Logistics Systems and Management, **15**(2), 288-303, 2013.
[13] E. Aljure, L. Busse, J. Hahn and S. Hayes, "Standardizing and improving the efficiency of Michigan Medicine's supply chain management of hospital supplies," 2017. http://umich.edu/~ioe481/ioe481_past_reports/17W01.pdf
[14] J.C. Chen, C.H. Cheng, P.B. Huang, K.J. Wang, C.J. Huang and T.C. Ting, "Warehouse management with lean and RFID application: a case study," he International Journal of Advanced Manufacturing Technology, **69**(1-4), 531-542, 2013.
[15] J. Benefield, "CPR For Lean Warehouses," Madison: Advantage Business Media. M., 2015. https://www.inddist.com/operations/article/13763532/cpr-for-lean-warehouses
[16] M. Bertolini, F. Bevilacqua, E. Ciarapica, and L. Postacchini "Business process reengineering of drugs storage and distribution: a case study" International Journal of Procurement Management, **8**(1/2), 44-65, 2015.
[17] D. Wua, J.E. Tepper, "Possibility of Inventory Pooling in China's public hospital and appraisal about its performance" Applied Mathematical Modelling, **39**(23-24), 7277-7290, 2015.
[18] M. Hammer, "Reengineering work: don't automate, obliterate," Harvard Business Review, **68**(4), 104-112, 1990.
[19] M. Motta, F. Lega, "Process re-engineering per le aziende sanitarie: una introduzione," Mecosan, **6** (23), 163-165, 1997.
[20] M. Al-Mashari, M. Zairi, "BRP implementation process: an analysis of key success and failure factors," Business process management journal **5**(1), 87-112, 1999.
[21] P.R. Mosca, "Il progetto quadrato latino," University of Genoa - Dipartimento di Ingegneria della Produzione - Gestione degli Impianti Industriali I
[22] D. Montgomery, Design and Analysis of Experiments (5th ed.). New York: Wiley, 2001, ISBN 0-7167-9657-0.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 263-266 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Design and Optimization of ZnO Nanostructured SAW-Based Ethylene Gas Sensor with Modified Electrode Orientation

Argha Sarkar[1,*], Padarthi Venkataramana[1], Nimmala Harathi[2], Thummuru Jyothsna[1], Neeruganti Vikram Teja[1]

[1]*Nanoelectronics Lab, Department of ECE, Sree Vidyanikethan Engineering College, 517102, India*

[2]*Nanoelectronics Lab, Sree Vidyanikethan Engineering College, 517102, India*

ARTICLE INFO

Article history:
Received: 01 November, 2019
Accepted: 19 January, 2019
Online: 27 January, 2020

Keywords:
Surface Acoustic Wave,
Hexagonal ZnO Nanorod,
Interdigitated Transducer,
Sensitivity.

ABSTRACT

Here the work approaches the detection of fruit maturity level. In addition, a gas sensor is proposed to obtain better sensitivity to know the level of ethylene gas released from matured fruit. Better sensitivity can be achieved by obtaining an enhanced active area, changing the intermediate layer thickness and the pitch value of the electrode. Hexagonal nanostructured zinc oxide is used as a sensing layer. Finally, frequency shift and a total displacement of the device are increased. The hexagonal structured ZnO nanorod-based surface acoustic wave gas sensor may be used to detect ethylene gas in fruit ripening sensor with enhanced sensitivity. This paper reports a simulation of Surface Acoustic Wave (SAW) wave propagation with respect to interdigitated transducer (IDT), piezoelectric material and sensing layers.

1. Introduction

It is becoming very important to design a highly sensitive fruit monitoring system to determine the storage life of fruit and quality. Based on the maturity level of fruit the decision may be taken regarding transportation, consumption, etc. The ripe fruits are very tough to handle as they become soft and squashy with an insipid smell and they cannot withstand transportation from one place to another. Hence it is important to monitor the state when they are mature. Skin color, hardness, size, and smell are the common parameters to get an idea about the fruit ripening. But these are basically assumption-based assessment, so it is havening an issue of reliability [1].

Ethylene (C_2H_4) is one of the organic gases which is formed in the ripening process of climacteric fruits. These fruits (eg. Mango, guava, banana, apricot, kiwi, palm, etc.) liberate C_2H_4 during ripening process [2-4]. Climacteric fruits pass through climacteric state after harvesting. A small dose of C_2H_4 in a controlled environment can simulate the ripening process of those climacteric fruits. So, if the maturity levels of these fruits are detected efficiently then they can be easily stored or transported after harvesting. On the other hand, non-climacteric fruits [5] emits a little amount of C_2H_4 and they are in a position to harvest only when completely ripened.

Herein, the work focuses on the development of surface acoustic wave-based ethylene gas sensor for detection of fruit ripening at different maturity stages. Detecting the C2H4 gas concentrations that has been liberated by fruits, may evaluate the maturity stage and the shelf span can be estimated [6]. In this study, a SAW-based sensor is designed and sensitivity is analyzed. Interigitated transducer (IDT), piezoelectric material and sensing layers are a major consideration while designing a SAW sensor. Pitch is one of the important parameters to determine the synchronous frequency. On the surface of the piezoelectric material, aluminum IDT generates acoustic wave from electrical signal by producing a mechanical force [7-8]. Adsorption of target gas molecules take place on the sensing material, which happens between the two IDTs. ZnO nanorod based SAW sensor is designed and wave properties are studied. Parameters like device length, height, intermediate layer thickness, and IDT thickness are optimized through simulation study to obtain enhanced sensitivity.

2. Sensor's Working Principle

The template SAW gas sensors are dominant from other sensors because of detection of physical, biological, and chemical quantities. It is only due to minimized insertion loss, increased sensitivity, and gas detection capability. Their fundamental operation is based on the changes in a propagating acoustic wave due to the target gas interaction, which depends on the the

*Corresponding Author: Argha Sarkar Email: argha15@gmail.com

propagation medium, mass density, electric–dielectric behavior and the elastic stiffness of the piezoelectric materials. Due to the piezoelectric effect, electrical signal is generated when the acoustic wave travels from the surface of the device substrate to an IDT. In SAW based gas sensor, a piezoelectric material plays a major role as a transducer. It transforms a signal from electrical to mechanical wave by utilizing the piezoelectric effect. Piezoelectric effect relates structural mechanics with the electrostatic field. The electric charge is created when a sinusoidal signal is applied to a piezoelectric material. The properties of a piezoelectric material rely on some important parameter like electric flux (D), resonating frequency (f_0), and stress (T). The functionalities of the SAW device are highly dependable on those parameters. To obtain the electrostatic fields (D) [9] from the equation 1, it is important to know about the parameters like stress vector (T), strain vector (S), electric field (E), dielectric matrix (e), piezoelectric matrix (ε) and elasticity matrix (c).

$$D = e^T S + \varepsilon E \tag{1}$$

From equation 1 it is cleared that it is required to know the structural mechanics of piezoelectric material [7,10]. Equation 2 is to find out the value of T.

$$T = cS - eE \tag{2}$$

Device efficiency is high when input signal frequency matches with resonating frequency. The resonating frequency of a SAW device is associated with propagation velocity (V_p) of the acoustic wave and the pitch value (p) (shown in equation 3 [11-12]).

$$f_0 = V_P / p \tag{3}$$

As per the equation 4, the central wavelength SAW sensor can easily be found from the electrode pitch (p) and finger width (h).

$$\lambda = 2(h + p) \tag{4}$$

IDT plays a vital role to obtain enhanced SAW sensor response. The lengths of the IDT, finger width, pitch are the important considerations. Bandwidth (B), acoustic wave frequencies are decided based on input IDT. For a specific operating frequency, the range of acoustic wave frequencies, can be enhanced by optimizing the number of pair of fingers (N) in IDT. Mathematically the bandwidth can be defined as [7,11],

$$B = \frac{2 f_0}{N} \tag{5}$$

If the number of pairs of finger in optimized the length of IDT (l_{IDT}) will also get changed.

$$B = \frac{2V_P}{NP} \tag{6}$$

$$B = 2\frac{2V_P}{l_{IDT}} \tag{7}$$

A sharp signal may be obtained by minimizing the bandwidth [11]. From equation 7 it can be said that bandwidth is inversely related to the length of the IDT. The sensing mechanism is mainly based on the adsorption phenomenon. When the sensor is placed in the target gas environment, target gas molecules are adsorbed on the surface of the sensing layer. It results in a change in the wave velocity. Because of the mass loading effect [13-14] there must be a frequency shift (given in equation 8) by which the detection can be performed.

$$\Delta f = (K_1 + K_2) f_0^2 \sigma - K_2 f_0^2 \sigma \frac{4\mu}{V_0^2}\left(\frac{\lambda + \mu}{\lambda + 2\mu}\right) \tag{8}$$

$$\Delta f = (K_1 + K_2) f_0^2 \sigma \tag{9}$$

(By neglecting the right part of equation 8)

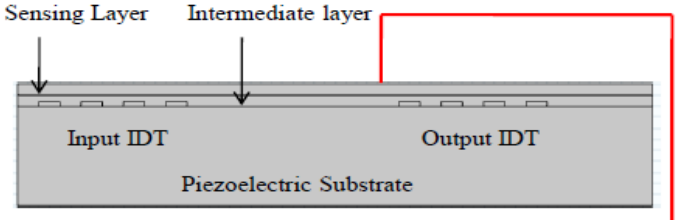
Where σ is the areal density of the layer after gas adsorption and λ, μ are signify Lame constant and shear modulus of the film respectively. If the value of Δf is low, there will be no significant change in frequency. It is not desired for a highly sensitive sensor. Δf is required to be significantly higher value. From equation 8 it can be said that $[K_2 f_0^2 \sigma \frac{4\mu}{V_0^2}(\frac{\lambda + \mu}{\lambda + 2\mu})]_{Min}$ gives the maximum frequency deviation. For a piezoelectric material $LiNbO_3$ and sensing layer ZnO, the value of shear modulus related expression becomes so small that $[K_2 f_0^2 \sigma \frac{4\mu}{V_0^2}(\frac{\lambda + \mu}{\lambda + 2\mu})]$ can be ignored [14].

The reduced expression is shown in equation 9.

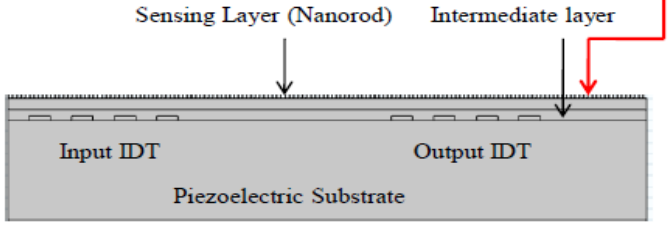
current designations.

3. Sensor Design and Simulations

Comsol Multiphysics 5.5 ® is considered to design and simulate the SAW-based gas sensor. Herein, initially simple ZnO based SAW gas sensor is considered and then it is embedded with ZnO nanorods of the surface of the ZnO sensing layer. Optimization of the delay-line two port configuration and applied frequency results in device performance. The important parameter of piezoelectric material lithium niobate is having a higher value of coupling coefficient and surface acoustic wave velocity. Lithium niobate is considered as a substrate material to optimize the device. The mass loading effect results after embedding MEMS module on the piezoelectric device. The frequency potential and displacement of the surface can be determined by using COMSOL Multiphysics. Sensing layer plays a vital part to get higher sensitivity. ZnO [15] is found to be an appropriate material towards gas sensing and it is also available in the list of built-in material in this simulation environment.



(a)



(b)

Figure. 1 Side view of the layout of (a) layered (b) Hexagonal nanostructured SAW gas sensor

The dimensions of the IDTs can be calculated by using equation 4 of 100 MHz signal frequency. Figure 1 a. shows the layout of the layered SAW gas sensor. Hexagonal nanorods are added on the nanolayered structure to obtain enhanced sensitivity. The sensitivity study is done based on the effect of hexagonal nanorods. In both cases, the simulation study is done with different values of device length, height, intermediate layer thickness, and IDT thickness. Here Figure. 1 (a) and (b) are the optimized design layouts.

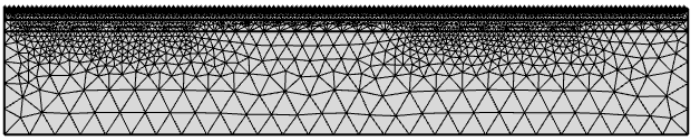


Figure. 2 2D meshing analysis of a SAW Sensor as the sensing material with nanorods

Where device length is 30 µm and its height is 4 µm. The thickness of the intermediate layer is 0.2 µm and ZnO layer is of 0.5 µm. The aluminum is taken as IDT material. The width of the IDTs is 1µm and height 0.1µm. To avoid the second order effect, electrodes are considered as a massless electrode. Figure. 2 represents the two dimensional cross sectional layout after doing the meshing. The node size is taken in such a way so that it becomes less than the wavelength and compensates the appropriate harmonics. Figure. 2 also ensures that the maximum number of nodes present is below the IDT location.

Where device length is 30 µm and its height is 4 µm. The thickness of the intermediate layer is 0.2 µm and ZnO layer is of 0.5 µm. The aluminum is taken as IDT material. The width of the IDTs is 1µm and height 0.1µm. To avoid the second order effect, electrodes are considered as a massless electrode. Figure. 2 represents the two dimensional cross sectional layout after doing the meshing. The node size is taken in such a way so that it becomes less than the wavelength and compensates the appropriate harmonics. Figure. 2 also ensures that the maximum number of nodes present is below the IDT location.

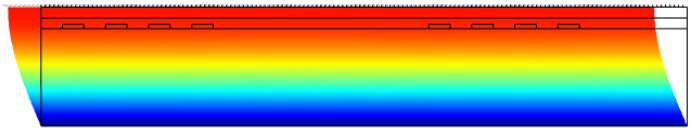


Figure. 3 Simulation of SAW Sensor as the Sensing Material with nanorods

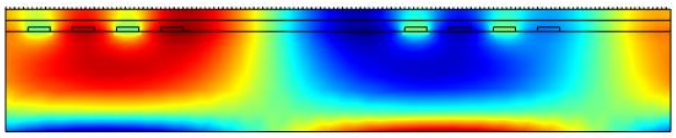


Figure. 4 Simulation of SAW Sensor as the Sensing Material with nanorods after exposing to gas

During the operation of the SAW device, a 150 Hz sinusoidal signal is given to the input IDT. It results in an electric field between the two opposite polarity IDT fingers. As the substrate is the piezoelectric material, due to the piezoelectric effect tensile and comprehensive strain will be created. Thus mechanical wave will appear. The wave will propagate through the surface of the material. Reflector and absorber are used to minimize the effect of insertion loss. When the analytes (ethylene) comes in contact with the sensing platform, sensing layer adsorbs the gas and results in increment in resonant frequency. Here, the sensing layer acts a vital role. By achieving higher surface to volume ratio of the nanostructured sensing material it can be expected that the adsorption of the target gas may be increased which in turn results in enhanced sensitivity. In this simulation study, the presence of ethylene gas is detected through the displacement in resonant frequency. Once the ethylene is detected with higher sensitivity, it will become easy to ensure the ripening state of the fruit.

4. Result and Analysis

After meshing and simulation (shown in Figure. 2 to Figure. 4) different colors have appeared. It is due to the applied voltage. More energy is confined above at the surface of the piezoelectric material [16]. Surface acoustic wave can be seen at the surface of the substrate yielding a noticeably change in propagation properties. The thickness of the sensing layer may effect on the sensitivity. A complete analogy of total displacement with respect to the device is shown in Figure. 5 and Figure. 6 upon exposure to ethylene gas.

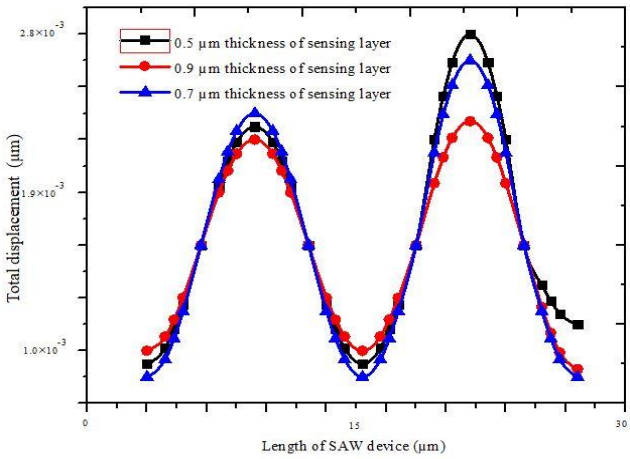


Figure. 5 Total displacements versus device length by varying the thickness of sensing layer

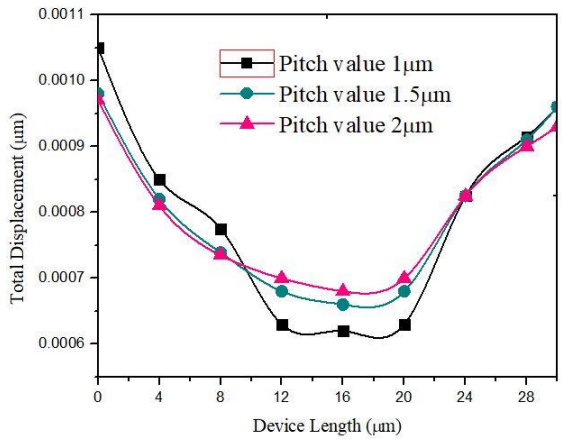


Figure. 6 Total displacements versus device length by varying the pitch value of IDT

Figure. 5 confirms that 0.5 μm thickness of the sensing layer provides increased displacement thus sensitivity. The dependency of sensitivity on pitch value is studied in Figure. 6. Aluminum is chosen as an IDT material because of its low sensitivity and low cost. Different pitch value of 1 μm, 1.5 μm and 2 μm are considered and it is found that 2 μm provides higher displacement of SAW at both input and output and output for a device length ranging from 8 to 20 microns. It results in enhanced sensitivity [17].

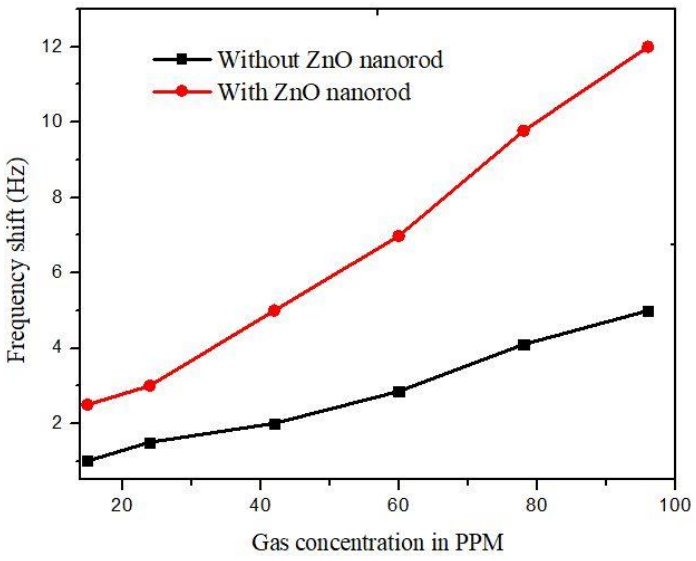


Figure. 7 Frequency shift due to gas exposure in presence and absence of ZnO nanorod

The frequency shift of the sensor is observed because of the different concentration of the target gas. The shift of frequency is present due to C_2H_4 gas release from fruit. The effects of ZnO nanorods on sensitivity are studied in Figure. 7. The presence of ZnO nanorod over the ZnO sensing layer exhibits better frequency shift compared to bared ZnO layered SAW sensor. Simulation results assure that ZnO nanorod based SAW sensor provides higher sensitivity. It is due to increased surface to volume ratio. Hexagonal nanorod provides increased surface to volume ratio i.e. more active region on the sensing layer which results more reaction with target analyte and increased displacement.

5. Conclusion

All An ethylene gas sensor to measure the fruit maturity is designed and simulated in COMSOL Multiphysics to study the results in resonant frequency by using ZnO sensing layer with proper electrode orientation. Furthermore, the ZnO nanorods are considered to obtain better sensitivity. A comparative study based ZnO layered and nanorod based sensors are performed in COMSOL Multiphysics. It reveals that nanorod based sensor provides enhanced frequency shift due to its higher surface to volume ratio in hexagonal structure. The displacement in resonant frequency is because of adsorption of the analyte by the sensing layer. Here future work includes the fabrication of ZnO nanorod and SAW sensor to verify the simulation results.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to acknowledge Sree Vidyanikethan Engineering College for infrastructural and laboratory support.

References

[1] T. Barboni, , M. Cannac, N. Chiaramonti, “Effect of cold storage and ozone treatment on physicochemical parameters, soluble sugars and organic acids in Actinidia deliciosa” Food chem., 121(4), 946-951, 2010. https://doi.org/10.1016/j.foodchem.2010.01.024

[2] S.P. Burg, E. A. Burg, “Relationship between ethylene production and ripening in bananas” Bot. Gaz., 126(3), 200-204, 1965. https://doi.org/10.1086/336320

[3] P. P. Iannetta, L. J. Laarhoven, Medina-Escobar, E. K. James, M.T. McManus, H.V. Davies, F.J. Harren, “Ethylene and carbon dioxide production by developing strawberries show a correlative pattern that is indicative of ripening climacteric fruit” Physiol. Plant., 127(2), 247-259, 2006. https://doi.org/10.1111/j.1399-3054.2006.00656.x

[4] C.S Barry, J. J. Giovannoni, “Ethylene and fruit ripening” J. Plant Growth Regul., 26(2), 143, 2007. https://doi.org/10.1007/s00344-007-9002-y

[5] S. O. Nelson, W. Guo, S. Trabelsi, Investigation of dielectric sensing for fruit quality determination. IEEE Sensors Applications Symposium, February, Atlanta, GA, USA. pp. 54-56, (2008). https://doi.org/10.1007/s00344-007-9002-y

[6] S. Janssen, K. Schmitt, , M. Blanke, M. L. Bauersfeld, J. Wöllenstein, W. Lang, “Ethylene detection in fruit supply chains” Phil. Trans. R. Soc. A: Mathematical, Physical and Engineering Sciences, 372(2017), 20130311, 2014. https://doi.org/10.1098/rsta.2013.0311

[7] J. Kirschner, “Surface Acoustic Wave Sensors (SAWS)” Micromechanical Syst., 1-11, 2010. https://doi.org/10.18689/mjnn-1000109

[8] M. Anand, ed., Study of tin oxide for hydrogen gas sensor applications. Thesis Scholar Commons University of South Florida, 2005.

[9] E. Comini, G. Faglia, G. Giorgio, Solid State Gas Sensing, Springer, 2009.

[10] Ballantine Jr, D.S., White, R.M., Martin, S.J., Ricco, A.J., Zellers, E.T., Frye, G.C. and Wohltjen, H., eds. Acoustic wave sensors: theory, design and physico-chemical applications. Elsevier, 1996.

[11] S. Capone, A. Forleo, L. Francioso, R. Rella, P. Siciliano, J. Spadavecchia, D. S. Presicce, A. M. Taurino, “Solid state gas sensors: state of the art and future activities” J Optoelectron Adv M, 5(5), 1335-1348, 2003. 10.1002/chin.200429283

[12] F. Nawaz, V. Jeoti, “SAW sensor read range limitations and perspectives” Wireless networks, 20(8), 2581-2587, 2014. https://doi.org/10.1007/s11276-014-0765-3

[13] M. N. Hasan, D. Acharjee , D. Kumar, S. Maity, "Simulation of low power heater for gas sensing application" Procedia Comput. Sci., 92 (2016), 213 – 221, 2016. 10.1016/j.procs.2016.07.348

[14] M. N. Hasan, S. Maity, A. Sarkar, C. T. Bhunia, D. Acharjee, A. M. Joseph, "Simulation and fabrication of SAW-based gas sensor with modified surface state of active layer and electrode orientation for enhanced H2 gas sensing" JEM, 46(2), 679-686, 2017. https://doi.org/10.1007/s11664-016-5128-7

[15] A. Sarkar, S. Maity, C.T. Bhunia, P.P. Sahu, "Responsivity optimization of methane gas sensor through the modification of hexagonal nanorod and reduction of defect states" Superlattices and Microstruct., 102, 459-469, 2017. https://doi.org/10.1016/j.spmi.2017.01.006

[16] G. Di Francia, B. V. Alfano, La Ferrara, "Conductometric gas nanosensors" J. Sens., 2009, pp 1-18, 2009. https://doi.org/10.1155/2009/659275

[17] G. Zhang, "Nanostructure-enhanced surface acoustic waves biosensor and its computational modeling" Journal of Sensors, 2009, 1-11, 2009. https://doi.org/10.1155/2009/215085

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 267-273 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Enabling 3D Heterogeneous Structures Towards Smart Chips: A Review

Cheng Li, Mengfu Di, Zijin Pan, Albert Wang*

Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521, USA

ARTICLE INFO

Article history:
Received: 23 November, 2019
Accepted: 19 January, 2019
Online: 27 January, 2020

Keywords:
Magnetic-cored inductor, nano crossbar array, graphene, NEM, switch, ESD protection, BEOL, metal wall, noise, non-binary

ABSTRACT

This review paper discusses recent research outcomes of fabricating three-dimensional (3D) heterogeneous structures in complementary metal-oxide-semiconductor (CMOS) integrated circuit (IC) technologies to enable smart future chips. The CMOS-based 3D heterogeneous structures demonstrated are vertical magnetic-cored inductors for radio-frequency (RF) ICs, through-back-end-of-line (BEOL) metal wall structures for global flying noise isolation, above-IC nano crossbar array electrostatic discharge (ESD) protection devices and graphene-based nano electromechanical (NEM) switch ESD protection structures, and individually controlled multiple-gated MOS field-effect transistors (MOSFET) for non-binary computing. These disruptive 3D heterogeneous structure concepts can potentially enable future smart IC chips in Si CMOS IC platforms.

1. Background

Since the birth of silicon monolithic integrated circuits (ICs) in 1959, geometrical scaling has been driving advances in silicon IC technologies, keeping on increasing transistor count at the Moore's Law pace until late 1990s [1]. The hallmark Si planar process was the technical foundation that has been guided by the National Technology Roadmap for Semiconductor (NTRS) through 1990s [2]. The laterally dimensional scaling has evolved into equivalent scaling, as guided by the International Technology Roadmap for Semiconductor (ITRS) until 2000 [3], where new materials and structural features, such as strained silicon and FinFET, were included. Later, 3D Power scaling kicked in to accommodate the ever-increasing demands for and emerging challenges of density, complexity and power as IC technologies continue to advance to sub-10nm nodes, as guided by ITRS 2.0 until 2015 [4]. Nevertheless, as 7nm CMOS is already in production today and 3nm CMOS is on the horizon, lateral shrinking is quickly approaching its end. Since the horizontal Si "land" is limited for increasingly complex ICs, going vertical to make Manhattan style Si-based "skyscrapers" becomes an obvious alternative. Meanwhile, future chips are not only bigger and more complicated, they ought to be smarter too, demanding for more device functionalities to handle more information diversity. Consequently, the ITRS 2.0 was brought to its end in 2015 [5] and the new Heterogeneous Integration Roadmap (HIR) was initiated to address the emerging technical needs and challenges in the new era of integrated circuits and systems [6-7]. The required functional diversification cannot not be delivered by conventional Si IC devices. Many novel devices and structures will be needed, which must be heterogeneously integrated into Si ICs in a 3D fashion, mainly based upon the CMOS platforms. To this end, substantial R&D efforts have been devoted to developing non-traditional 3D heterogeneous devices and structures to enhance functionality and improve performing characteristics of conventional Si-based CMOS ICs. For example, Nanoelectromechanical (NEM) relay structure in CMOS backend was recently reported to realize reconfigurable interconnects to enable complex and hard-reconfigurable ICs [8]. This review paper, an extended version of our paper presented at IEEE International Conference on IC Design and Technology [9], provides an overview of our recent research outcomes of developing various disruptive 3D heterogeneous structures to enable smart future CMOS chips.

2. Vertical Magnetic-Cored Inductors

Proliferation of wireless communications have changed our society and the foundation is RF IC technology. The ultimate design goal is to achieve RF system-on-a-chip (SoC) to ensure higher system performance and reliability while reducing costs. The main technical barrier for making RF SoCs is lack of super compact, multiple-GHz, high-performance IC inductors in CMOS. Traditional IC inductors using metal spirals are too large, typically bigger than a bonding pad, which makes single-chip RF SoC impractical. The dream for RF IC designers is to have high-

*Corresponding Author: Albert Wang, Email: aw@ece.ucr.edu

performance transistor-sized inductors in CMOS. For years, various magnetic-enhanced inductor structures have been reported, utilizing lateral magnetic films or complex MEMS structures to reduce magnetic losses. These magnetic inductors are impractical to CMOS ICs. As a solution, we proposed a novel stacked-via vertical-magnetic-cored inductor structure made in the back-end-of-line (BEOL) of CMOS [10-17]. To understand our disruptive concept, imagine to shrink a perfect discrete solenoid inductor and plugged it vertically into the CMOS BEOL stack, like a "needle", which forms a transistor-sized vertical-magnetic-cored inductor in CMOS BEOL stacks. As the inductor size shrinks, electrical loss increases due to increased series resistance of metal wires, which however, can be compensated by reducing magnetic loss by use of a closed-loop magnetic circuit. Hence, the overall inductor performance does not suffer from inductor shrinking. Figure 1 depicts the concept of the new vertical- magnetic-cored inductor. Taking full use of CMOS BEOL structures, the vertical magnetic bar can be fabricated by locally replacing the common metal vias in a CMOS BEOL stack with suitable magnetic materials, hence called a stacked-via magnetic-cored inductor. One critical specification for the magnetic-cored inductor is to achieve high performance (e.g., Q-factor and inductance density, L-density) at high frequencies (>1GHz). Substantial theoretical and experimental work has been conducted to prototype the new inductor devices and improve the inductor performance [10-19]. Various magnetic materials compositions, materials synthesis techniques, magnetic core structures and process steps were explored. In our initial work, ferrite thin film materials, e.g., Ni0.4Zn0.4Cu0.2Fe2O4, Y2.8Bi0.2Fe5O12 and Co7ZrO9, were used as the magnetic media. A CMOS-compatible process module was developed using a spin-coating method to make single-layer ferrite magnetic enhanced inductors. As depicted in Figure 2, measurements showed encouraging inductance improvement over the non-magnetic references (R1/R2), from 5.8% to 38% up to 9GHz. To enhance magnetic materials for multiple-GHz operations, we further engineered various nano particle powders for the magnetic cores, including Co_2Z-type ($Ba_3Co_2Fe_{24}O_{41}$) and Ni–Zn–Cu ($Ni_{1-x-y}Zn_xCu_yFe_2O_4$) series. We found that the Ni–Zn–Cu ($Ni_{1-x-y}Zn_xCu_yFe_2O_4$) composites achieved the best reported performance up to ~10GHz. We then developed a post-CMOS process flow to fabricate the vertical magnetic cores for RF inductors as shown in Figure 3. The new vertical magnetic-cored inductors were fabricated in a foundry 180nm 6-metal CMOS technology. Shown in Figure 4, the new magnetic-cored inductors achieved the then- best-reported inductor performance, i.e., Q_{max}·f_{max}~L-density characteristics: ~220% improvement in Q-factor to up to 20GHz and the highest reported L-density of ~700nH/mm^2 at $Q_{max}f_{max}$ of ~4GHz. A new 3D modeling technique was developed to accurately simulate any arbitrary conductor-magnet structures to guide the design and optimization of vertical magnetic-cored inductors [18, 19]. The new stacked-via magnetic-cored structure can be used to fabricate other types of compact high-performance inductive devices, which may not only reduce RF IC chip size, but also potentially allow new chip architecture. For example, a bank of compact magnetic-cored transformers can be used to realize simplified power amplifier (PA) architecture to address the power back-off problem. The new magnetic-cored inductors were applied to design of LC-VCO ICs in CMOS [20, 21]. Figure 5 shows that in a 1.36–1.86GHz VCO fabricated in a 180nm RF CMOS, it achieves a high FOM factor of ~202dBc/Hz and a reduced phase noises of -121 dBc/Hz at 100-kHz [12]. In a 2.22-2.92GHz VCO fabricated in a 180nm SOI CMOS, it achieves phase noise reduction for from -106.97dBc/Hz to -113.49dBc/Hz [21].

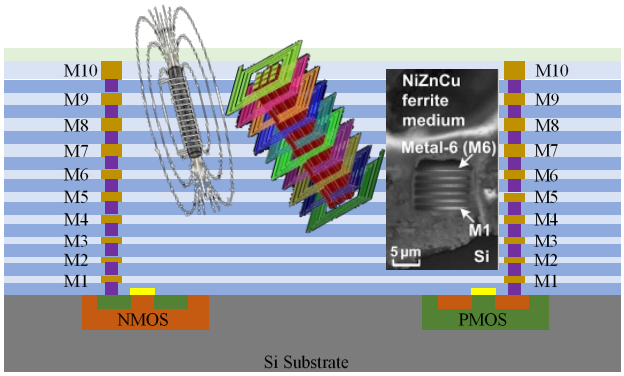


Figure 1 A conceptual view of new vertical magnetic cored inductor in CMOS platform. Insets from Left to Right: a discrete solenoid, simulated vertical magnetic-cored inductor and fabricated vertical magnetic-cored inductor.

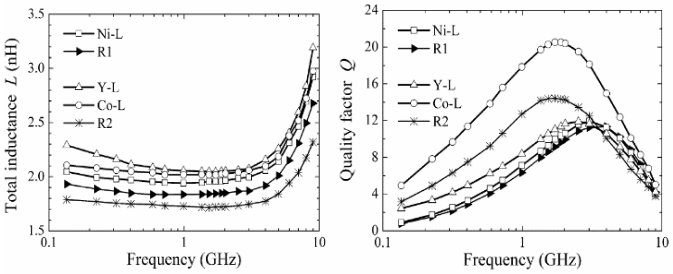


Figure 2 Measurement of initial single-layer magnetic-enhanced inductor shows improvement in inductance and Q-factor: Ni-L *versus* R1 and Y-L/Co-L *versus* R2 [10].

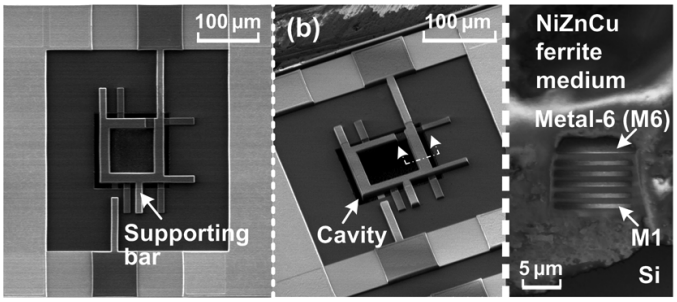


Figure 3 SEM photos for samples of fabricated vertical magnetic-cored inductors in a 180nm 6-metal CMOS.

3. Through-BEOL Metal Wall Noise Isolation

Interferences, i.e., noise or crosstalk, affect IC performance. As technology shrinking and chip scale increasing continue, noise/crosstalk isolation becomes extremely important and challenging, particularly for large and complex SoC chips. For years, substantial research has been devoted to develop various noise-isolation techniques, resulted in many isolation methods, such as, guide rings, deep trenches, etc. These noise-isolation structures are all inside Si substrates, hence, they only work for in-Si noise isolation.

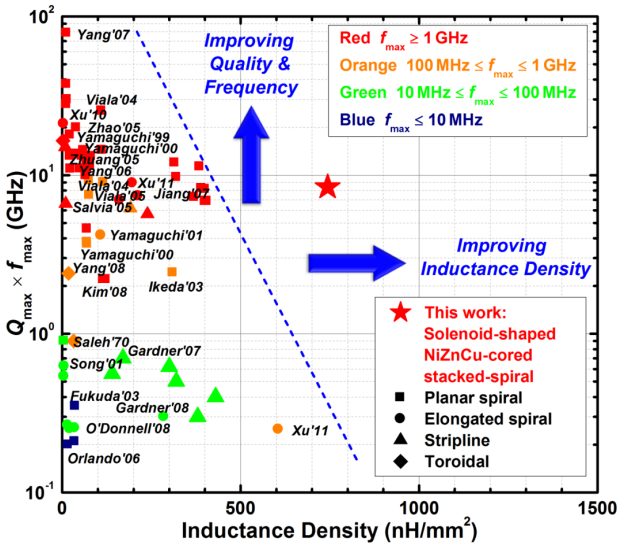


Figure 4 Comparison shows that the vertical magnetic-cored inductor (the star) out-performs the reported state-of-the-art of various magnetic-enhanced inductors.

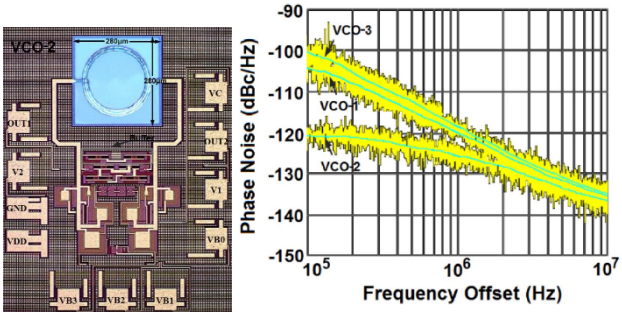


Figure 5 Measurement shows that VCO2 using magnetic-cored inductor has better phase noise performance compared to VCO1/VCO3 using conventional inductors [20].

For complex ICs at advanced nodes, noise coupling mainly occurs in the BEOL stacks through the extremely complex and conductive metal interconnects mesh. This new type of in-BEOL noise coupling is defined as "flying noise", which is global in nature across the whole chip, and unfortunately, cannot be suppressed by any conventional in-Si noise-isolation techniques. To address this emerging global flying noise challenge, we proposed a novel through-BEOL metal wall flying noise isolation structure as illustrated in Figure 6 [22, 23]. In principle, it is a (near) closed-loop metal wall structure that blocks transmission of electromagnetic waves, hence, noise-isolates circuit blocks residing inside the through-BEOL metal cages. From EM theory, the through-BEOL meta walls can be optimized in terms of metal materials and wall physical designs to block the noise propagation, therefore, eliminating the global flying crosstalk in the BEOL stacks. This flying noise isolation concept was validated experimentally in CMOS. Figure 7 shows the first experiment where a simple amplifier circuit was designed in a foundry 180nm SOI CMOS. A post-CMOS process module was developed to etch a deep trench into the CMOS BEOL stack, which was filled with silver powder to form the in-BEOL metal wall structure. In measurement, it was observed that the troublesome third-order intermodulation interferers were suppressed by about 9dB (Figure 7) [22].

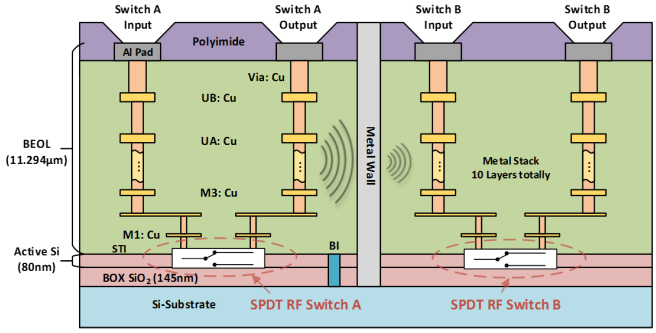


Figure 6 A conceptual X-section view of the new through-BEOL metal wall flying crosstalk isolation structure in SOI CMOS.

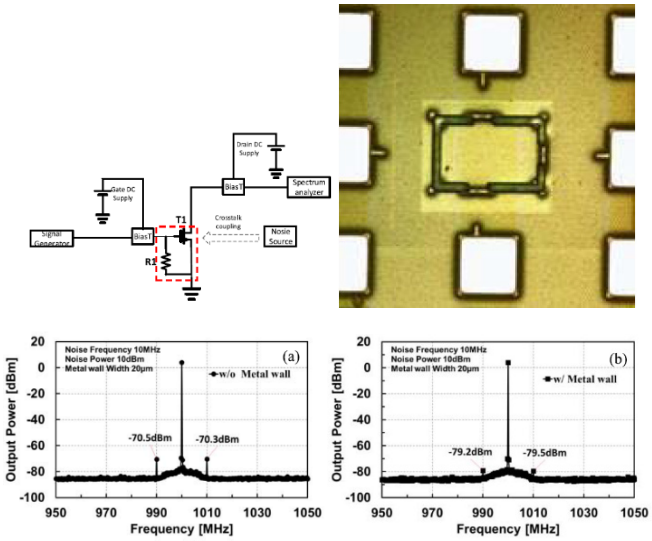


Figure 7 An amplifier circuit with the new in-BEOL metal wall can reduce the IM3 interferers by 9dB in measurement [22].

Figure 8 presents a second prototype design, which are SPDT RF switch circuits fabricated in a foundry 45nm SOI CMOS [23]. The design splits include SPDT switches with and without in-BEOL metal walls that were made of silver nano powders.

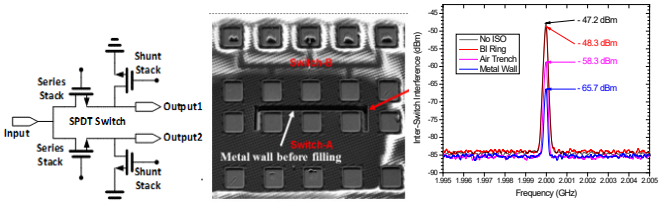


Figure 8 SPDT RF switches fabricated in a 45nm SOI CMOS utilizes the new through-BEOL metal wall to successfully block the global flying crosstalk by 98% in linear scale [23].

Measurement shows that the in-BEOL metal wall structure reduces the inter-switch flying crosstalk by 18.5dB, almost entirely blocked the global flying crosstalk (~98.6% in linear scale). The experiments confirmed that our new through-BEOL metal wall flying noise isolation technique is efficient in blocking the above-Si global flying crosstalk.

4. Above-IC Nano-Crossbar ESD Protection

Electrostatic discharge (ESD) failure is the most devastating reliability problem to ICs. On-chip ESD protection is required for all ICs and electronic systems [24]. As IC process technologies continue scaling and chip complexity rapidly increases, on-chip ESD protection design emerges as a major IC reliability design challenge. Since the initial ESD awareness in late 1960s, ESD protection has been relied on in-Si PN-junction-based active devices to form low-impedance conduction channels to discharge fast ESD transients, hence, provide ESD protection for ICs. Unfortunately, the conventional ESD protection structures inevitably introduce parasitic effects inherent to the PN junctions, such as ESD-induced parasitic capacitance, noises and leakage current, which seriously affect IC performance, particularly for high-frequency broad band RF ICs and high-speed ICs. On the other hand, on-chip ESD protection structures consume large Si die area in order to handle the significant ESD-related energy, and the typically irregular ESD device shapes makes IC physical layout design truly challenging. These ESD protection design problems are becoming unbearable to advanced ICs at sub-28nm nodes. It is hence imperative to develop disruptive ESD protection solutions for future chips that are not tied to in-Si PN junctions [25]. To address such emerging ESD protection design challenges, we devised two revolutionary above-IC non-PN type ESD protection concepts and device structures. Figure 9 depicts the first concept of an above-IC nano-crossbar array ESD protection structure [26-28]. In principle, this new device is a two-terminal phase-changing switch structure comprising two electrodes and a phase-changing insulator in between. The insulator is made of specially synthesized nano materials that is electrically OFF in normal IC operations. Under ESD stressing, the strong incident electrostatic force causes phase change of the nano insulator, making it conducting, hence turns ON the nano crossbar switch structure to discharge the ESD transients. When an ESD event is over, the switch will return to its OFF state. The nano crossbar array switch

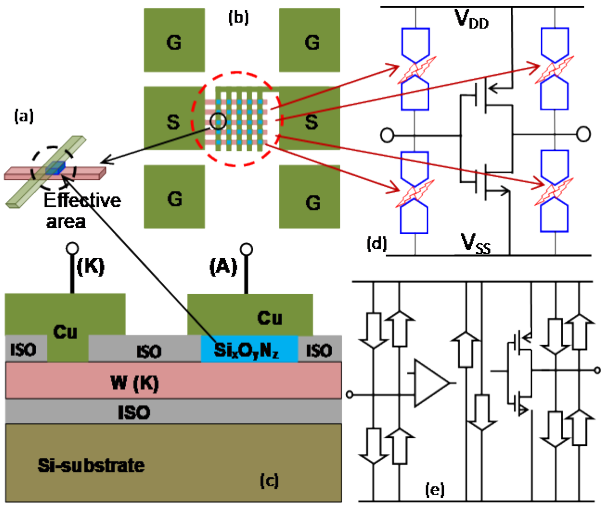


Figure 9 A conceptual illustration for the new above-IC nano crossbar array ESD protection structure in CMOS.

is built in the CMOS BEOL stack above the Si substrate and is not a conventional PN junction device. It is an above-IC ESD switch structure. While each crossbar node is a nano device, a large nano crossbar array can be designed to handle large ESD energy collectively. This novel nano crossbar array ESD switch structure was experimentally validated using a post-CMOS process module developed. Figure 10 presents the measured transient ESD I-V curve for a single-node crossbar ESD switch by transmission-line pulsing (TLP) ESD testing, showing the desired symmetrical two-directional ESD discharging I-V behavior. Figure 11 shows the measured I-C curves for a nano crossbar array ESD switch where multiple ESD triggering effect can be observed. Measurement shows ultrafast ESD response time of ~100pS, negligible leakage current of I_{leak}~2pA (Figure 12), tunable ESD triggering voltage V_{t1} of 1.26~28.3V (Figure 13) and ultrahigh ESD protection capability of ~267V/μm^2. Since the new nano crossbar ESD protection structure is above the Si substrate, it does not consume extra Si die area. This novel above-IC nano crossbar array ESD switch structure can potentially resolve all challenging problems associated with the conventional in-Si PN-based ESD protection structures.

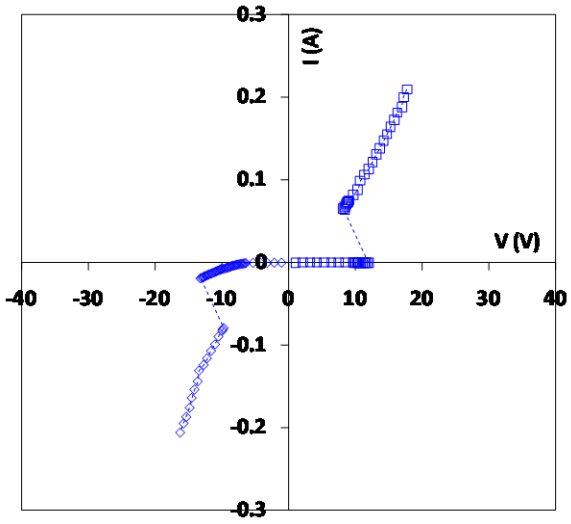


Figure 10 TLP testing shows symmetric ESD discharging I-V curve for a single-node nano crossbar ESD protection structure.

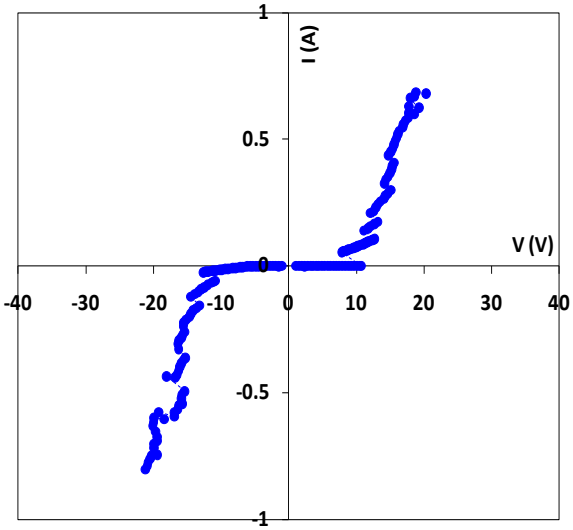


Figure 11 TLP testing shows symmetric ESD discharging I-V curve for a sample nano crossbar array ESD protection structure.

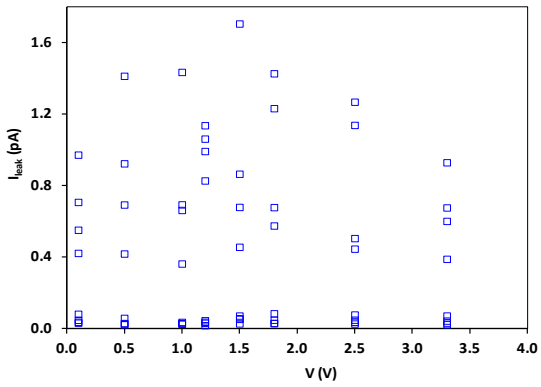


Figure 12 Measurement shows ultralow leakage currents, due to the non-PN ESD structure, for sample nano crossbar ESD protection structures.

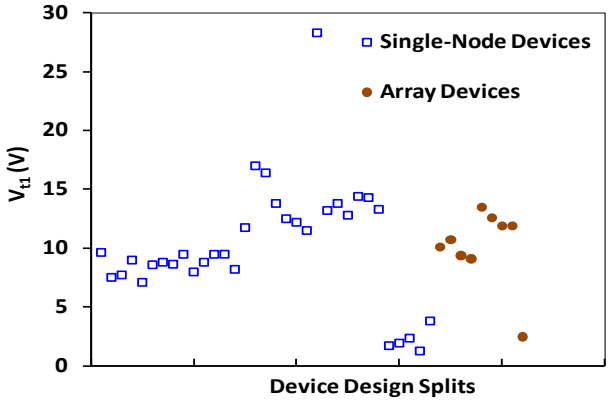


Figure 13 Measurement shows tunable ESD triggering voltage for sample nano crossbar ESD protection structures.

5. Above-IC Graphene NEMS ESD Switch

Figure 14 shows the second novel above-IC ESD protection structure, which is a graphene-based NEMS (gNEMS) switch ESD protection device fabricated in the CMOS BEOL stack [29, 30]. This new graphene gNEMS ESD switch is a two-terminal device, entirely different from conventional in-Si PN-based ESD protection devices. A gNEMS has a graphene membrane over a cavity. The two electrodes are the graphene membrane on top (cathode, or, K) and the bottom of the cavity (anode, or, A). Unlike any PN-based electronic devices, the gNEMS device is a mechanical switch, which remains electrically OFF in normal IC operations. When an ESD transient appears, the electrostatic force pulls the graphene membrane downward to touch the bottom conductor, hence turns on the gNEMS device to discharge the ESD surge. When the ESD transient is over, the graphene film will bounce back and turns the gNEMS OFF again. Since gNEMS is a mechanical switch, ideally, it has no leakage current and almost no parasitic capacitance. Since it is built in the CMOS BEOL stack, above the IC substrate, gNEMS devices will not consume extra Si die area. Therefore, the gNEMS ESD switch is an ideal ESD protection structure. Since graphene film is ultralight and has very large Young's module, the gNEMS switching speed can be extremely fast. Figure 15 presents the measured transient ESD discharging I-V behaviors for a sample gNEMS switch, showing the desired dual-directional and symmetrical I-V characteristic. Measurement also shows negligible leakage current of ~3pA, superfast ESD response time of ~200ps and ultrahigh ESD protection capability of ~1500V/μm^2, which is many orders of magnitude higher than a silicon-controlled rectifier (SCR) ESD protection structure that is normally considered the most robust Si ESD protection structure. It is encouraging that the gNEMS ESD switch can be potentially ideal ESD protection solution for future ICs.

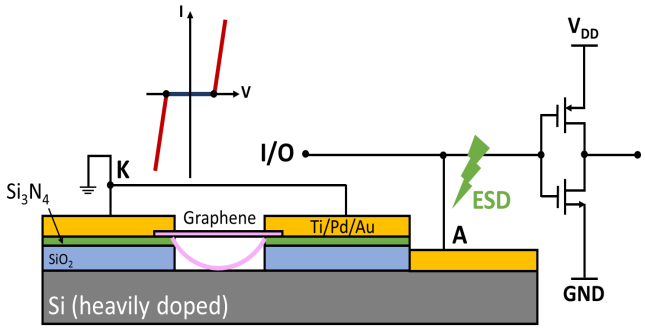


Figure 14 A conceptual illustration for the new above-IC gNEMS ESD switch device in CMOS.

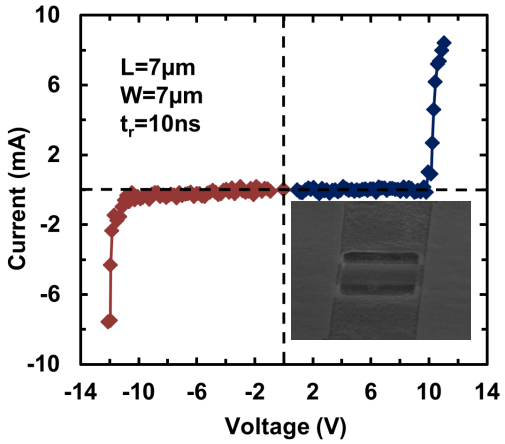


Figure 15 TLP testing shows a dual-directional ESD discharging I-V curve for a sample gNEMS ESD switch, where the inset is a SEM image for gNEMS.

6. Multiple-Gated MOSFETs

Our information and internet era rely on binary data processing, including computing and storage, which was enabled by transistors and CMOS ICs. The base-2 binary numbering system is the foundation for most information processing systems, whose capacity is certainly limited, i.e., 2^n (n is bit number in a byte) caps binary the information system capacity. The binary numbering system is natural to CMOS ICs because a MOSFET has two conduction states, OFF (0) and ON (1) corresponding to I_D = 0 or not. Imagine a base-M numbering system with M>2, where $M^n > 2^n$ results in a much bigger system capacity for a given byte size. Research has been done to explore multiple-state-value devices, mainly pursuing for unique conducting materials to realize many-status logics. As a potential solution, we proposed a novel individually-controlled multiple-gated MOSFET structure. Figure 16 depicts such a device concept where a MOSFET has four individually controlled gates over the same conduction channel. By selectively biasing the four gates, individually or collectively, this MOSFET will deliver multiple conduction states, hence, forms

a non-binary M^n numbering system [31, 32]. For different channel conduction states, I_D will be different for the MOSFET. By designing the 4-gated MOSFET, 5<M<16 may be obtained. There are several advantages for this base-M information MOSFET structure and numbering system. First, its information system capacity increases significantly from the binary system. Second, by choosing M=10 through device design, a base-10 (i.e., 10^n) system may be realized, which not only dramatically increases the system capacity, but also makes computing algorithm and system architecture more natural to human instinct. Third, this 4-gated MOSFET can be fabricated using classic CMOS IC technologies through 3D heterogeneous integration.

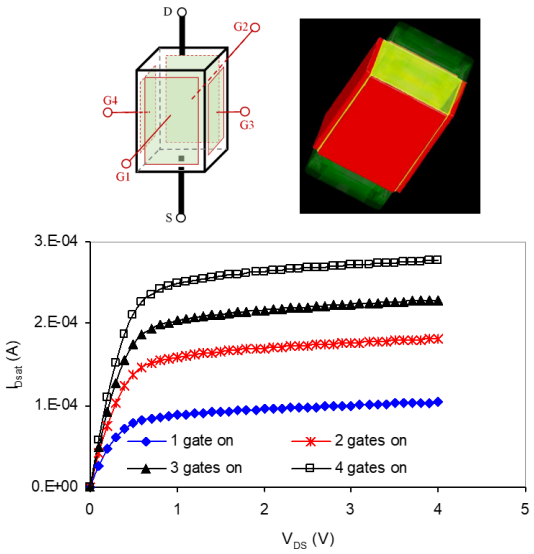


Figure 16 Conceptual illustration for the novel multiple-gated MOSFET.

7. Summary

The past six decades have witnessed the unprecedent advances of the Si-based IC technologies, which has completely changed human life. As we are rapidly moving from the information technology (IT) era to the internet-of-everything (IoET) age, it calls for disruptive IC technologies to facilitate the emerging needs for information sensing, computing, storage and communications, which simply cannot be supported by the hallmark Si planar IC process technologies. To break the deadlock of space, complexity and performance of ICs, one viable answer to the grand technical challenge is to develop smart future chips that can integrate more device functionalities, in a 3D fashion, to improve operating characteristics of IC chips. 3D heterogeneous structures and heterogeneous integration are the potential solutions to continuous advances of Si-based CMOS ICs beyond the Moore's Law. This paper reviews some novel 3D heterogeneous structures that we developed recently to address the emerging IC design challenges including: compact inductors with vertical magnetic cores, in-BEOL global flying noise isolation structures, on-chip ESD protection and multi-gated transistors for non-binary numbering systems. These new 3D heterogenous structure concepts, while being exemplar, are truly disruptive and potentially enablers for future smart chips.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors acknowledge National Science Foundation for partial supports of the related research.

References

[1] G. E. Moore, "Cramming more components onto integrated circuits", Electronics, pp. 114-117, April 19, 1965. DOI: 10.1109/N-SSC.2006.4785860
[2] SIA National Technology Roadmap for Semiconductor (NTRS)
[3] SIA International Technology Roadmap for Semiconductor (ITRS)
[4] SIA International Technology Roadmap for Semiconductor 2.0 (ITRS 2.0), 2015 Edition.
[5] ITRS wensite, http://www.itrs2.net/
[6] Heterogeneous Integration Roadmap (HIR), 2019 Edition.
[7] Heterogeneous Integration Roadmap (HIR) website, https://eps.ieee.org/technology/heterogeneous-integration-roadmap/2019-edition.html
[8] Z. A. Ye, S. Almeida, M. Rusch, A. Perlas, W. Zhang, U. Sikder, J. Jeon, V. Stojanović, and T.-J. K. Liu, "Demonstration of 50-mV Digital Integrated Circuits with Microelectromechanical Relays", in IEEE IEDM Tech Digest, 2018. DOI: 10.1109/IEDM.2018.8614663
[9] C. Li, F. Zhang, M. Di, Z. Pan and A. Wang, "Developing 3D Heterogeneous Structures for Future Chips", in Proc. IEEE International Conference on IC Design and Technology (ICICDT), pp.48-51, 2019. DOI: 10.1109/ICICDT.2019.8790942
[10] C. Yang, F. Liu, T. Ren, L. Liu and A. Wang, "On-Chip Integrated Inductors with Ferrite Thin-films for RF IC", in IEEE IEDM Digest, pp.225-228, 2006. DOI: 10.1109/IEDM.2006.347003
[11] J. Zhan, C. Yang, X. Wang, Q. Fang, Z.T. Shi, Y. Yang, T.-L. Ren, A. Wang, Y.H. Cheng, L.-T. Liu, "Stacked-spiral RF inductors with vertical nanoparticle magnetic core for radio-frequency integrated circuits in CMOS", Sensors & Actuators: A. Physical, V195, pp. 231-238, June 2013. http://dx.doi.org/10.1016/j.sna.2013.02.030
[12] Z. Ni, J. Zhan, Q. Fang, X. Wang, Z. Shi, Y. Yang, T.-L. Ren, A. Wang, Y. Cheng, J. Gao, X. Li, and C. Yang, "Design and Analysis of Vertical Nanoparticles- Magnetic-Cored Inductors for RF ICs", IEEE Trans. Electron Devices, **60**(4), pp. 1427-1435, April 2013. DOI: 10.1109/TED.2013.2245418
[13] J. Zhan, C. Yang, X. Wang, F. Zhang, T.L. Ren, A. Wang, Y. Yang, L. T. Liu, L. W. Yang and Z. X. Yue, "Stacked-Spiral RF Inductor with Vertical Nano-Powder-Magnetic-Core in CMOS", IEEE Microwave and Wireless Components Letters, **22**(1), pp.29-31, January 2012. DOI: 10.1109/LMWC.2011.2176474
[14] C. Yang, F. Liu, X. Wang, J. Zhan, A. Wang, T-L. Ren, L-T. Liu, H. Long, Z. Wu, X. Li, "Investigation of on-Chip Soft-Ferrite-Integrated Inductors for RF ICs — Part I: Design and Simulation", IEEE Trans. Electron Devices, **56**(12), pp.3133-3140, December 2009. DOI: 10.1109/TED.2009.2033328
[15] C. Yang, F. Liu, X. Wang, J. Zhan, A. Wang, T-L. Ren, L-T. Liu, H. Long, Z. Wu, X. Li,, "Investigation of on-Chip Soft-Ferrite-Integrated Inductors for RF ICs — Part II: Experiments", IEEE Trans. Electron Devices, **56**(12), pp. 3141-3148, December 2009. DOI: 10.1109/TED.2009.2033413
[16] C. Yang, F. Liu, T. Ren, L. Liu, G. Chen, X. Guan, A. Wang and H. Feng, "Ferrite-Integrated on-Chip Inductors for RF ICs ", IEEE Electron Device Letters, **28**(7), pp.652-655, July 2007. DOI: 10.1109/LED.2007.900299
[17] C. Yang, F. Liu, T-L. Ren, L-T. Liu, G. Chen, X-K. Guan, A. Wang and Z-X. Yue, "Ferrite-Partially-Filled on-Chip RF Inductor Fabricated Using Low-Temperature Nano-Powder-Mixed-Photoresist Filling Technique for Standard CMOS", in IEEE IEDM Tech. Digest, pp1038-1040, 2007. DOI: 10.1109/IEDM.2006.347003
[18] H. Long, Z. Feng, H. Feng and A. Wang, "A novel accurate PEEC-based 3D modeling technique for RF devices of arbitrary conductor-magnet structure", Microwave and Optical Technology Letters, 38(3), Wiley & Sons, pp237-240, August 2003.
[19] H. Long, Z. Feng, H. Feng, A. Wang, T. Ren, J. Bao, F. Liu, C. Yang, X. Zhang, "A New Modeling Technique for Simulating 3D Arbitrary Conductor-Magnet Structures for RFIC Applications", IEEE Tans. Electron Devices, **52**(7), pp. 1354-1363, July 2005. DOI: 10.1109/TED.2005.850662
[20] H. Cai, Y. Yang, N. Qi, X. Chen, H. Tian, Z. Song, Y. Xu, C-J. Zhou, J. Zhan, A. Wang, B. Chi and T-L. Ren, "A 2.7-mW 1.36–1.86-GHz LC-VCO With a FOM of 202 dBc/Hz Enabled by a 26%-Size-Reduced Nano-Particle-Magnetic-Enhanced Inductor", IEEE Trans. Microwave Theory and Techniques, **62**(5), pp. 1221-1228, May 2014. DOI: 10.1109/TMTT.2014.2312886
[21] R. Ma, F. Lu, Q. Chen, C. Wang, F. Liu , W. Zou and A. Wang, "A 2.22–2.92GHz LC-VCO Demonstrated with An Integrated Magnetic-Enhanced

Inductor in 180nm SOI CMOS", in Proc. IEEE RFIC Symp., pp.110-113, 2016. DOI: 10.1109/RFIC.2016.7508263

[22] C. Wang, Q. Chen, F. Lu, C. Li, F. Zhang and A. Wang, "Blocking Flying Crosstalk in BEOL Validated in Antenna Switches in 45nm SOI CMOS", IEEE Microwave and Wireless Components Letters, **28**(11), pp.1005-1007, November 2018. DOI: 10.1109/LMWC.2018.2873197

[23] F. Lu, Q. Chen, C. Wang, F. Zhang, C. Li, R. Ma, X. S. Wang and A. Wang, "In-Die Through-BEOL Metal Wall for Noise Isolation in 180nm FD-SOI CMOS", IEEE Electron Device Letters, **38**(5), pp.630-632, May 2017. DOI: 10.1109/LED.2017.2682819

[24] A. Wang, On-Chip ESD Protection for Integrated Circuits, Kluwer, Boston, 2002.

[25] A. Wang, H. Feng, R. Zhan, H. Xie, G. Chen, Q. Wu, X. Guan, Z. Wang and C. Zhang, "A review on RF ESD protection design", IEEE Trans. Electron Devices, **52**(7), pp.1304-1311, July 2005. DOI: 10.1109/TED.2005.850652

[26] L. Wang, X. Wang, Z. Shi, R. Ma, J. Liu, Z. Dong, C. Zhang, L. Lin, H. Zhao, L. Zhang, Y. Cheng, R. Huang and A. Wang, "Dual-Directional Nano Crossbar Array ESD Protection Structures", IEEE Electron Device Letters, **34**(1), pp.111-113, January 2013. DOI: 10.1109/LED.2012.2222337

[27] L. Lin, L. Zhang, X. Wang, J. Liu, H. Zhao, H. Tang, Q. Fang, Z. Shi, R. Huang, Y. Cheng and A. Wang,, "Novel Nanophase-Switching ESD Protection", IEEE Electron Device Letters, **32**(3), pp.378-380, March 2011. **DOI:** 10.1109/LED.2010.2099100

[28] L. Zhang, R. Huang, D. Gao, D. Wu, Y. Kuang, P. Tang, W. Ding, Y. Wang and A. Wang, "Unipolar Resistive Switch Based on Silicon Monoxide Realized by CMOS Technology", IEEE Electron Device Letters, **30**(8), pp.870-872 ,August 2009. DOI: 10.1109/LED.2009.2024650

[29] R. Ma, Q. Chen, W. Zhang, F. Lu, C. Wang, A. Wang, Y. H. Xie and H. Tang, "A Dual-Polarity Graphene NEMS Switch ESD Protection Structure", IEEE Electron Device Letters, **37**(5), pp. 674-676, May 2016. DOI: 10.1109/LED.2016.2544343

[30] W. Zhang, R. Ma, Q. Chen, M. Xia, J. Ng, A. Wang and Y-H. Xie, "The electro-mechanical responses of suspended graphene ribbons for electrostatic discharge applications", Applied Physics Letter, **108**, 153103 (2016). *http://dx.doi.org/10.1063/1.4946007*

[31] A. Wang, "Novel 3D 4-Gate MOSFET and 5-ary Logic for Hi-density Hi-speed Memory", a proposal to SRC, Sept. 2003.

[32] C. Li, F. Zhang, M. Di, Z. Pan and A. Wang, "Advances in 3D Heterogeneous Structures and Integration for Future ICs", in Proc. IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), 2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 1, 274-284 (2020)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

NAO Humanoid Robot Obstacle Avoidance Using Monocular Camera

Walaa Gouda[*,1],Randa Jabeur Ben Chikha[2]

[1] *Computer and Engineering Networks Department, Jouf University, Saudi Arabia*

[2] *Computer Science Department, Jouf University, Saudi Arabia*

ARTICLE INFO

Article history:
Received: 07 December, 2019
Accepted: 08 January, 2020
Online: 10 February, 2020

Keywords:
Motion planning
Humanoid
Object recognition

ABSTRACT

In this paper, we present an experimental approach that allows a humanoid robot to effectively plan and execute whole body motions, like climbing obstacles and straight stairs up or down, besides jumping over obstacles using only on-board sensing. Reliable and accurate motion sequence for humanoid employed in complex indoor environments is a necessity for high-level robotics tasks. Using the robot's own kinematics will construct complex dynamic motions. A series of actions to prevent the object from being performed on the basis of the identified object from the database of the robot, obtained using the robot's own monocular camera. As shown in real world experiments beside simulation using NAO H25 humanoid, the robot can effectively perform whole body movements in cluttered, multilevel environments containing items of various shapes and dimensions.

1 Introduction

Humanoid robots have become a common research platform because they are considered the future of robotics because of their creativity. Humanoid robots, however, are fragile mechanical robotic systems and the main challenge is to maintain the balance of the robot [1]. Human like design and locomotion require complex motions to be performed by humanoids. Humanoids are well adapted for human-designed mobile manipulation activities such as walking, reaching various types of terrain, moving in complex environments such as environments with stairs and/or narrow passages, navigating in cluttered environments without colliding with any barriers, etc. Such capabilities would make humanoid robots perfect assistants to humans,such as disaster management or housekeeping [2]–[5].

Humanoid service robots need to deal with a variety of objects. For example, avoiding objects by turning away from them, stepping over,onto or down an obstacle, climbing up or down stairs. For any humanoid robot, these are considered challenging tasks. Humanoids usually perform motion instructions improperly [2, 3, 6]. That is because they possess rough odometry estimates; they might slip on the ground surface depending on the ground friction, as well as the frequency of its joint back-lashing. In fact, their light-weighted and small sensors are influenced by noise inherently. All these circumstances can lead to uncertain estimates of pose and/or inaccurate execution of movement [7, 8]. Nonetheless, there are other explanations that justify why in practical applications humanoids are not used often. For example, humanoids are expensive as they are produced in small quantities and consist of complex hardware parts [3].

Many researchers use navigation algorithms that use wheel robots rather than legs, but the disadvantage of this model is that it does not respect all the collision avoidance capabilities of humanoids; therefore, more appropriate methods are needed to navigate in cluttered and multilevel circumstances [3, 9]–[11].

At first, Humanoid research has concentrated on some variables such as basic walking, but current systems are now becoming more knotted. Most humanoids already have full body control prototypes and advanced sensors such as sound, laser, stereo vision, and touch sensor systems. That provide the circumstances required and important to deal with complex issues foe instance grasping and walking. Movement planning is a positive way of dealing with complex issues, because planning enables the versatility of meeting different criteria. The design of complex dynamic motions is accomplished only by *robot kinematics*, which transforms from the common space where the kinematic chains are described into the Cartesian space where the robot's manipulators travel and

[*]Corresponding Author:Walaa Gouda, Computer and Engineering Networks Dept., Jouf University, Saudi Arabia, wgaly@ju.edu.sa

vice versa [3, 12]–[14].

As it can be used to plan and execute movements, robot kinematics is important, it is split into *forward* and *inverse* kinematics. Forward kinematics corresponds to using the kinematic equations of the robot to calculate the position of the end effector from the specified parameter values [12, 15, 16]. On the other hand, reverse cinematics corresponds to using the kinematics equations of a robot to determine the joint parameters that provide the target position of the end effector. That is clear why kinematics is necessary in any aspect of complex movement design [12, 14, 16, 17]. The relationship between inverse and forward kinematics is shown in Figure 1.

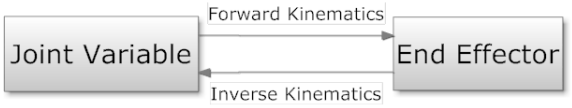


Figure 1: Representation of inverse and forward kinematics

Whole body motions that fulfill a variety of constraints are required for Humanoids to perform complex motions activities, where the robot has to maintain its balance, avoid self-collisions and collisions with environmental obstacles. However, consideration must be given to the ability of humanoids to move over/ down/ onto objects and maneuver in multi-level environments. All these restrictions and the high degree of freedom of the humanoid robot make the whole body motion planning a challenge [2]. The main goal of *whole body balance motion* is to make consistent motions and adapt the robot's behavior to new situations [18].

Based on our Knowledge, human-like movements depend on the characteristics and the objectives of given tasks. A possible solution to the evolution of human-like motions or attitudes is to investigate and solve a given task with software of a human robot. In order to reduce these human efforts and perform more complex tasks, many researchers have studied another approach based on learning theories. Demonstration robot programming, also known as learning imitation, tedious manual programming seemed to automate the manipulation of robots [19, 20, 14].

For many supported tasks to be performed autonomously and safely, a robot must be aware of the task restrictions and must prepare and carry out motions that take into account these constraints while avoiding obstacles [19, 20]. Such motion planners, however, typically require manually programming the task constraints, which requires a domain knowledge programmer. On the other hand, methods focused on learning from demonstrations are highly effective in automatically learning task restrictions and controllers from people demonstrations without programming knowledge and experience [20].

Humanoid motion design has been examined thoroughly in recent years. For example, the approach provided by [7] allows a NAO humanoid to climb spiral staircases efficiently, fitted with a 2D laser sensor and a monocular camera..

Although [9] proposed a method for locating NAO humanoids using only on-board sensing in undefined complicated indoor environments. Approach *(Nishiwaki et al.*[21] allows NAO to ascend single steps after manually aligning the robot in front of them without any sensory information to detect the stairs. While in [22], the authors provides a three-step action plan for ascending staircases for HRP-2. Furthermore in [23], the authors proposed a technique for climbing stairs using efficient image processing techniques with single camera fixed at the top of the robot at a height of 60 cm.

In addition, *Burgetet al.* [4] introduce an approach to the whole body movement design by manipulating interactive items such as doors and drawers. Their experiments with a NAO unlocking a drawer, a lock, and carrying up an object showed their power to solve complicated planning problems. *Gouda et al.* [24] proposes another approach to entire body motion planning for humanoids using only on-board sensing. NAO humanoid has been used to verify the sequence of actions posed in order to avoid obstacles, step over acts and step up or down.

Thus in [3], the researchers developed a motion named T-step, which helps the robot to move over actions as well as to parameterize steps on or down actions. The paper presented by *Gouda et al.*[1] suggested a NAO humanoids method for climbing up stairs as well as jumping up or down obstacles placed on the ground.

Measurement of distance for robots is very important as the robot needs to go to the exact position to perform additional tasks, such as playing with a ball, observing the house etc. The authors in [25] used four different measurement methods to compare the results. Such methods include the live Asus Xtion Pro, sonar, NAO-owned cameras for achieving stereo vision, and NDI for obtaining the 3D coordinate and ground truth as well. They conclude by selecting an infrared tool such as Asus Xtion Pro Live for short range calculation of distance. They also deduced that both sonar and infrared can be easily used to obtain the depth information without much intervention, but more studies needs to be done with stereo vision.

This paper is a major overhaul and extension of earlier [24, 1] conference versions. We have developed a NAO humanoid movement system that generates solutions that meet all the necessary constraints and apply reverse kinematics to joint chains.

The designed framework allows the robot to intelligently perform whole body balancing action sequences, involving walking over, up or down obstacles , as well as climbing up or down straight stairs in a 3D environment, shown in Figure 2. Based solely on the on-board sensors and joint encoders of the robot, we built an effective whole body movement strategy that performs safe motions to maneuver robustly in challenging scenes with obstacles on the ground, as shown in Figure 2. Our approach uses a single camera to determine the appropriate movement consisting of a series of actions based on the observed obstacle. As shown in the practical experiments with a NAO humanoid H25 as well as a series of simulation experiments using Webots (a simulation applications for programming, modeling and simulating robots [26, 27]), our system leads to vigorous whole body

movements in cluttered multilevel environments comprising barriers of various sizes and shapes.

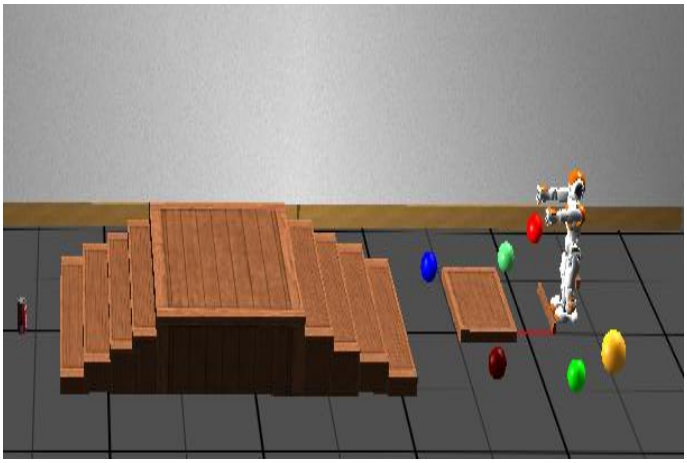


Figure 2: The simulated environment similar to the real world environment

2 Methodology

The obstacle avoidance device problems presented in this paper are generally applied to the humanoid robot of NAO. NAO robot (shown in Figure 3) is the world's premier research and learning humanoid. It is the perfect platform for all forms of science, technology, engineering and mathematics. [28, 29].

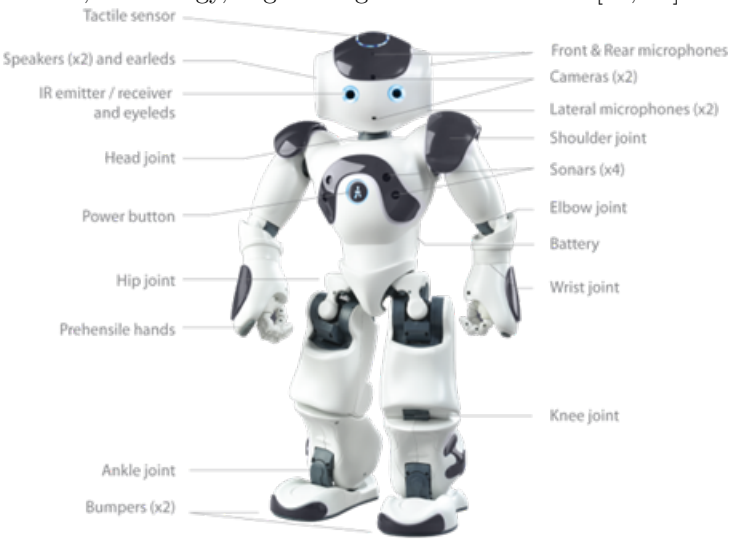


Figure 3: Aldebaran NAO H25 [18]

NAO humanoid robot is an integrated, programmable, medium-sized humanoid experimental platform with five kinematic chains (head, two arms, two legs) developed by Aldebaran Robotics in Paris, France. The NAO project began in 2004. NAO officially replaced Sony's quadruped AIBO robot in the RoboCup in August 2007. NAO has produced many prototypes and several models over the past few years [18, 14, 30].

Table 1 summarized NAO Humanoid Robot's technical specification information, where NAO H25 is about 58 cm in height; 5.2 kg weighs and includes a programming framework called NaoQi that offers a user-friendly interface for controlling sensors on the NAO and sending control commands to achieve higher robot behaviour. The NAO robot used has 25 degrees of freedom (DOF), each joint is fitted with location sensors, it has 11 DOF for its lower parts and 14 DOF for its upper parts [18, 31, 32].

Table 1: Technical specification of NAO Humanoid Robot

NAO's Descriptions	NAO's Specifications
NAO's Height	58 cm
NAO's Weight	5.2 kg
NAO's Autonomy	90 min const. walk
NAO's Degree of freedom	25 DOF
NAO's CPU	X86 AMD Geode 500 MHz
NAO's OS (Built-in)	Linux
NAO's Compatible OS	Mac OS, Windows, Linux
NAO's Programming languages	C, C++, .Net, Urbi, Python
NAO's Vision	Two CMOS 640x480 cameras
NAO's Connectivity	WiFi, ethernet

NAO incorporates a variety of sensors that can be used to identify the face and form with two similar video cameras on the forehead. Nonetheless, the two cameras do not overlap, as shown in Figure 4, And each time only one is active and the view can be switched almost instantly from one to the other.

Both cameras provide a 640x480 pixel video with a speed of 30 frames per second, which can be used to distinguish visual items such as goals and balls [18]. Four sonars on the chest (2 receivers and 2 receivers) allow NAO to detect obstacles in front of it. Furthermore, NAO has a wealthy torso inertial system with a 2-axis gyroscope and a 3-axis accelerometer that can provide real-time information about its instantaneous body movements. Two bumpers at each foot's tip are basic off/on switches and can provide obstacle information on foot collisions. There is a set of strength-sensitive resistors for each foot that support the forces applied to the feet, while encoders on all stepper motors record the real measurements of all joints for each moment. [14, 30].

Eventually, a GEODE 500 MHz board with 512 flash memory and optional extension via a USB bus is the CPU located in the robot's head. If necessary, a Wifi connection connects NAO to any local network or other NAOs. NAO is powered by Lithium Polymer batteries with time ranging from 45 minutes to 4 hours depending on their operation [18, 31].

In general, the robot should be completely symmetrical, but it is noteworthy, according to the manufacturer, that some joints on the left side differ from the corresponding joints on the right side [14, 30]. In addition, certain joints appear to be able to shift within a wide range, the robot's hardware controller forbids access to the limits of these

ranges due to possible NAO shell collisions [14]. Kinematics is very useful for developers of NAO software because it can be used to plan and execute these complex movements [14, 6, 30].

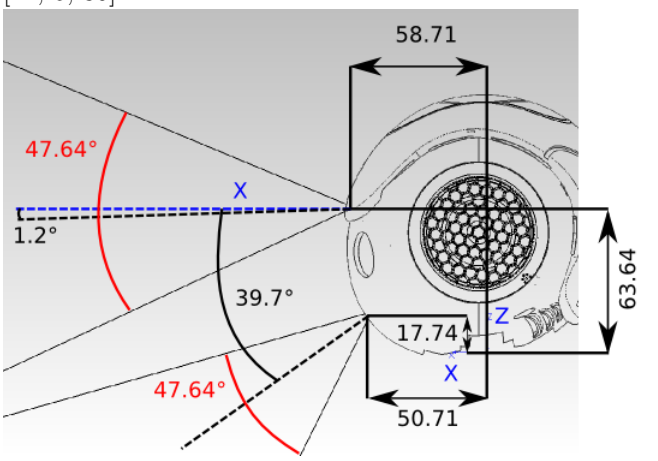


Figure 4: Cameras Location [18].

NAO's geometric model gives the location of the effector ($X = [P_x, P_y, P_z, P_{wx}, P_{xwy}, P_{wz}]$) relative to an absolute space according to all the common positions ($q = [q_1, ..., q_n]$).

$$X = f(q) \tag{1}$$

The direct kinematic model is presented in equation (2), which is the time derivative of equation (1).

$$\dot{X} = \frac{\delta}{\delta t} f(q)\ \dot{q} = J(q)\ \dot{q} \tag{2}$$

Where $J(q)$ is the Jacobian matrix. It is necessary to check the end effector and deduct the joint position, so that the inverse kinematic model shown in equation (3) is needed.

$$\dot{q} = J^{-1}\dot{X} \tag{3}$$

In many cases, J is not explicitly inversible directly (matrix not square), this problem is solved mathematically using Moore-Penrose pseudoinverse [18].

Aldebaran Robotics includes shared values in the documentation of the robot [18]. For each link / joint, the center of mass is represented by a point in the three-dimensional space that assumes the joint's zero pose.

The obstacle avoidance model problems suggested in this work are generally applied to the humanoid robot of NAO. Since NAO H25 has 25 DOF, it can perform many complex motions including walking, kicking a ball, going up and down obstacles, etc. [14, 6].NAO's walk uses a simple dynamic model (Linear Inverse Pendulum) which is stabilized by feedback from its joint sensors, which makes the walk stable against minor disturbances and absorbs frontal and lateral torso oscillations [18].
The swing foot can be positioned on the front and 16 cm on the side at most 8 cm and the peak elevation is 4 cm using the walking controller given. The robot's feet are about 16 cm x 9 cm in size. From these statistics, it is clear that NAO can not perform complex motions such as avoiding obstacles (excluding turning itself around) using the standard motion controller as shown in Figure 5 [18, 3, 14].

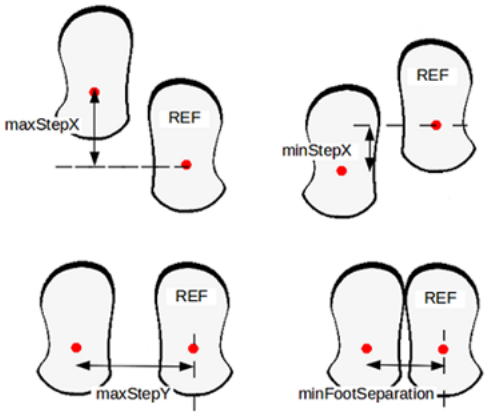


Figure 5: Clip with maximum outreach [18].

Only the on-board sensor of the robot, the lower camera, is used in this paper. Another limitation for NAO is its camera, as the two cameras of NAO do not overlap and only one of them is active each time and the view can be switched almost instantly from one to the other [18]. That means that it is not possible to form a 3D image, i.e. details about the distance between the obstacle and the robot.

In order to avoid obstacles, NAO must identify objects in the environment if they encounter some, so that it must first know specific objects by using the Choregraphic Vision Monitor [31] (Choreography is a graphical device created for NAO humanoids by Aldebaran Robotics).

Algorithm 1: Recognize objects

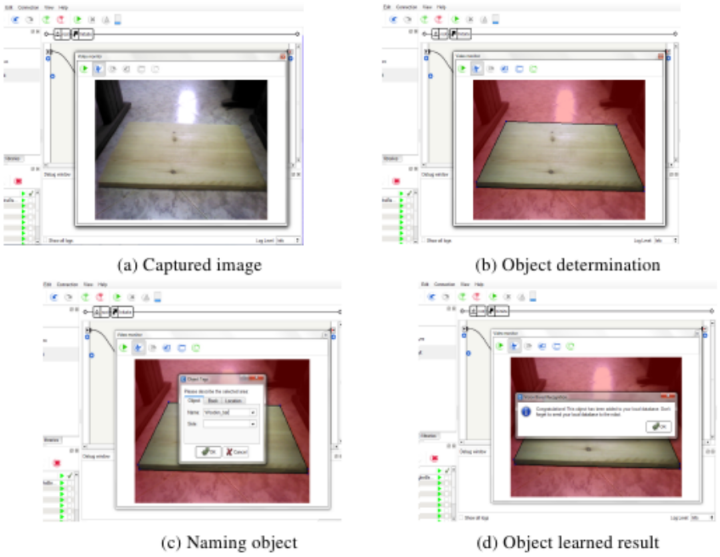
```
while There exist some objects to learn do
    capture image using NAO's cameras ;
    draw the contour of the learned object manually ;
    give a name and other details to the object ;
end
storing learned images in NAO's database ;
```

Object learning is shown in Figure 6 and described in Algorithm 1. The image should be obtained using NAO's cameras as shown in Figure 6a, then the contour of the learning object is drawn, segment by segment, manually as shown in Figure 6b. After that, in order to differentiate between different objects, details about the object like the name is entered as shown in Figure 6c.

Then a message appears to show the learning process's success (see Figure 6d). Photos are stored in NAO's database after learning all the necessary items. Once all photos are placed in NAO's database, during its active deployment, NAO will be able to perform object recognition. The method of recognition is based on visual key points detection and not on the object's external form, so NAO can only identify

objects that have been previously studied.

The recognition method is partly resilient to distance, varying from half to twice the distance used for learning, and angles up to an inclination of 50° for something observed facing the camera, rotation and light conditions [18]. Each key point observed in the current image is matched only with one of the key points learned in the database. So, if the score is too small to choose between two items, none of them will be correlated with the key point. This means that the object can not be learned twice, because learning the same area of an object twice reduces its detection rate as the score is the same for both objects and the object is not recognized [18].



(a) Captured image (b) Object determination

(c) Naming object (d) Object learned result

Figure 6: Object learning phase

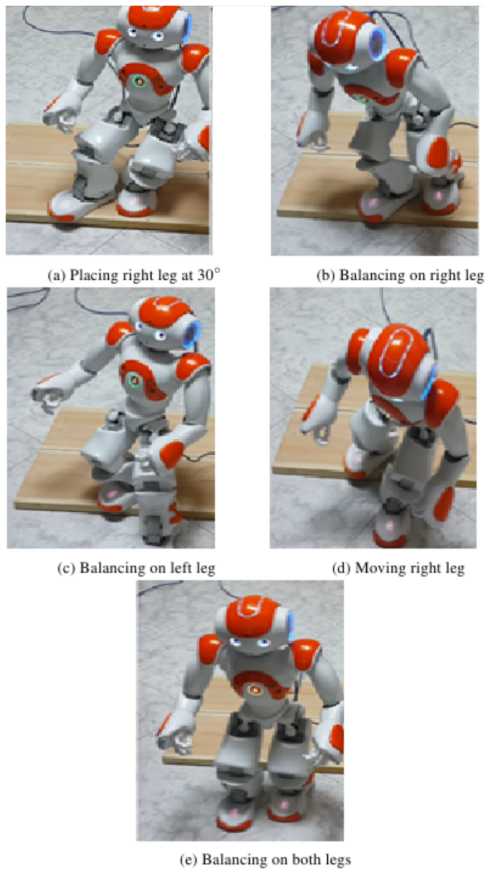
2.1 Motion design

As shown in Figure 5, it is obvious that NAO can not use its typical motion controller to perform complex motions (other than walking) such as moving over/onto/ down obstacles. In order to allow the robot to overcome these limitations, a *kinesthetic teaching* is applied. Here, the humanoid NAO H25 is programmed using the choreograph [31] and the python programming language. We implement a special movement design (described in Algorithm 2), inspired by [3] and extends the work in [24, 1].

This design allows the robot to walk over iteratively as well as to climb up or down obstacles (including straight staircases) depending on the shape of the obstacle. The right leg's foot (L_R) is positioned in the intended motion at an angle of 30°, which is the basis for the other acts. The robot then shifts his balance to that leg, (L_R), (the leg with the angle 30°) and moves freely the other leg, left leg (L_L). Then the balance is shifted to L_L, then L_R moves freely next to L_L and then the balancing is performed on both legs as shown in Figure 7. The development of such complex dynamic movements can only be accomplished by using the robot's own kinematics [2, 12].

Algorithm 2: Actuate motion

```
go to initial position;
place right foot at angle 30°;
move balance to right foot;
let left foot move freely;
if obstacle recognized is a long bar then
    move left foot forward;
    set the foot height to zero, i.e., put it on the ground;
else if obstacle recognized is a wide bar then
    move left foot forward;
    if robot on the ground then
        set the foot height to 2cm, i.e., onto the bar;
    else [robot on the bar]
        set the foot height to -2cm, i.e., down the bar;
    end
else the recognized object is stairs
    move left foot forward;
    if robot facing stairs up then
        set the foot height to 4cm, i.e., climbing up stairs;
    else [robot on the bar]
        set the foot height to -4cm, i.e., climbing down stairs;
    end
end
move balance to left foot;
let right foot move freely;
move right foot forward beside the left foot;
```



(a) Placing right leg at 30° (b) Balancing on right leg

(c) Balancing on left leg (d) Moving right leg

(e) Balancing on both legs

Figure 7: Designed motion step

The motive for using this series of motions is to manipulate, when stepping forward, the greater lateral foot displacement. From such a position, the robot can perform step over

movements to overcome obstacles with an elevation of up to 5 cm which can not be done with the standard motion controller.

The sequence of motions for ascending and descending stairs or wooden bar is similar to the sequence of motion for stepping over actions except for the positioning the swing foot, as it is positioned closer to the placed foot but at a different height where the height is changed using reverse kinematics depending on the recognized item.

The movements designed for full body stability use NAO's original kinematics to directly control the effectors in Cartesian space using an inverse kinematics approach to solving. The model *Generalized Inverse Kinematics* is used; it addresses cartesian and joint power, stability, redundancy, and role priorities.

This formulation takes into account all the joints of the robot in a single problem. The obtained motion guarantees several specified conditions, such as balancing, holding down one foot, etc. The developed motion system's capabilities are then demonstrated in a series of simulation experiments using Webots (a simulation platform for modeling, programming and simulating robots [26]) for the NAO robot, as well as real-world experiments with the real NAO H25 robot.

The main differences between the motion built in this work and the motions defined by *Maier et al.* [3] are in the angle of placement of the robot's foot. Because our model movement helps the robot to position its foot at a 30° angle, This allows the robot to achieve the equilibrium in a shorter time and more securely than the motion given by *Maier et al.* [3], while *Maier et al.* [3] equals 90° to the robot's positioning angle.

Our work further expands the work outlined in [1] as it executes the entire scenario starting with walking in search of a target, including avoiding obstacles by ascending straight stairs, climbing up stairs as well as walking over, on or down obstacles until the goal is reached.

3 Experimental Evaluation

The developed framework allows the robot to robustly execute entire body balancing action sequences, including stepping over and ascending or descending obstacles, as well as ascending or descending straight stairs in a 3D environment, shown in Figure 2. Based solely on the onboard sensors and joint encoders of the robot, we built an effective whole body movement strategy that performs safe movements to maneuver robustly in demanding scenes with obstacles on the ground as shown in Figure 2.
Our method uses monocular camera to determine the appropriate motion consisting of a series of actions based on the detected obstacle. As shown in practical experiments with a NAO humanoid H25 as well as a series of simulation experiments using Webots, the framework results in efficient motions for the whole body in congested multilevel environments with artifacts of various shapes and sizes.

The scenario consists of a world in which the humanoid has to navigate through arbitrarily cluttered obstacles (see Figure 2). The robot is going to navigate the environment until it reaches the goal of recognizing the cane.

The target is far from the robot's reachable space at the start of the experiment. The robot starts to walk in quest of its target to avoid obstacles. Once the obstacle is in front of the robot, the entire body movement must be carried out in order to prevent the robot from colliding with it.

It's worth noting that if the robot gets lost and encounters items that it hasn't learned instead of locating the previously learned objects or target, it's going to notify a message that notifies it is lost and stop walking.

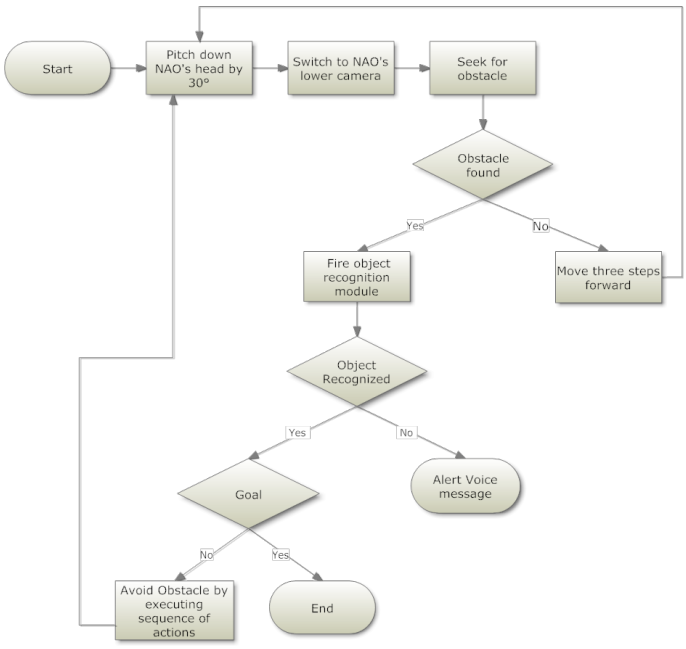


Figure 8: The main process of our experimentation on NAO H25

Figure 8 illustrates the main process that describes our approach. The robot must move three steps forward in the experiments presented, then stop moving and pitch his head down at an angle of 30° and switch to lower camera in his head to look for obstacles on the ground in front of his feet. The object recognition module will be activated until an obstacle is detected; then the robot must take three more steps forward. In the case of an obstacle being recognized, the robot compares the object being recognized with the objects being observed and stored in its database.

If it is the robot's goal, it must stop moving and alert a voice message that notifies it reaches its goal; otherwise it will execute stable whole body motions to deal with the

recognized object and then move three steps forward and so on until the goal is achieved. If the robot does not know the object, it triggers a voice message notifying that it is approaching an object that it has not previously known or seen and then stops moving.

Figure 9 shows the operation of choreographic [31] identification of items [31], the top right of the photo shows a video monitor showing what the robot actually sees. It is clear that through the robot's vision recognition module, the robot can recognize the long wooden bar previously taught and saved in NAO's database. The vision recognition module returns the name of the recognized object to the next stage when a match occurs to take an action based on the object. Figure 10 Is the action module expansion shown in Figure 9, where the robot has to perform step by step sequence of motions to overcome this barrier when the wooden bar is recognized.

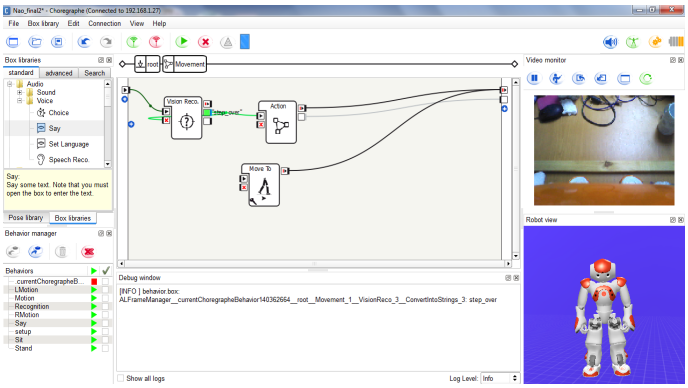


Figure 9: Wooden bar is recognized

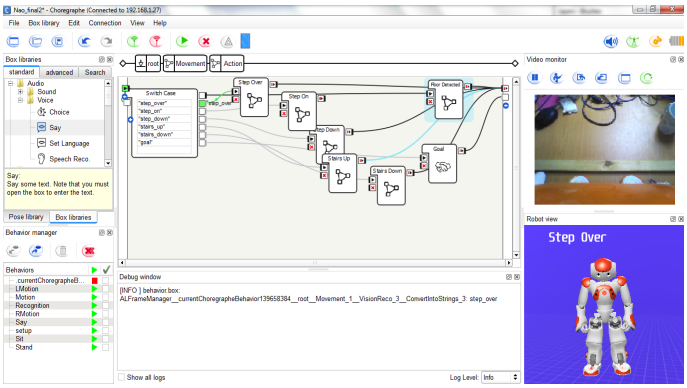


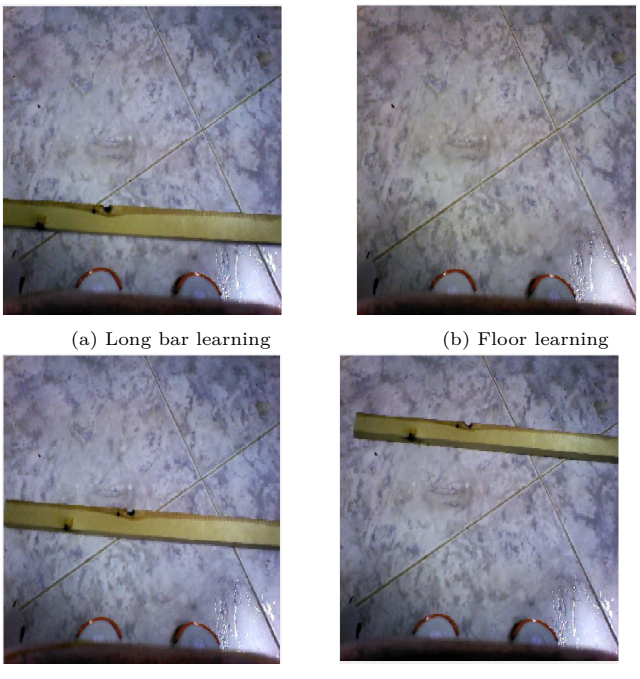
Figure 10: Step over actions are executed

The key challenge is to accurately determine the distance from NAO to the obstacle. As mentioned earlier, the NAO camera does not provide the distance between the robot and the obstacle. The robot's vision recognition module is also partly robust to distance the robot from the target, as it ranges from half to twice the distance used to learn [18]. It is also based on the identification of visual key points, not on the object's external shape, which means that if the robot recognizes the wooden bar, it will perform step-over activities regardless of the distance between them.

If the distance between his feet and the target is not sufficient, the robot may struggle to overcome the obstacle. The distance between the robot and the obstacle is not known because the camera of the robot has a limitation in providing depth information. If the obstacle is placed below the correct margin of the robot, the robot will strike the obstacle while moving its leg, resulting in a shift in the angle of the foot, disrupting its balance and the robot will fall. A further situation if the obstacle is placed at a distance greater than a suitable distance, the robot can put his swing foot on the bar, which will also cause a disturbance in his balance and fall.

A simple solution to this problem is to allow the robot to learn the image of the wooden bar on the floor at a distance near his head as shown in Figure 11a, and learn the image of the floor in the learning phase near the robot's feet as shown Figure 11b.

If the wooden bar is located at a large distance in front of the robot as shown in Figure 11d and Figure 11c, both the floor image and the long wooden bar image will be recognized by the vision recognition module. But as the vision recognition module gives high visual key points to the floor over the long wooden bar, it will return the floor as the recognized object. In this situation, the robot pushes three steps forward, until the long wooden bar's visual key points are higher than the floor's, which means the robot is now close to the wooden bar and must be avoided. In deciding the distance between NAO and the barrier, this solution proves to be very efficient.



(a) Long bar learning (b) Floor learning

(c) Floor Detection (d) Floor Detection 2

Figure 11: Learning floor image

The experiments are conducted on: (a) the robot stepping over a 40 cm wide wooden bar, 3.5 cm high and 2 cm deep; see Figure 12, (b) Step on a wooden bar 40 cm wide, 2 cm

high and 40 cm deep; see Figure 14, and (c) move down to the ground from that bar as shown in Figure 15. Figure 16's first raw shows the process of climbing up a single step of a 40 cm wide, 4 cm high and 20 cm deep straight staircase. The second row shows the remaining four steps of the staircase climbing up. Similarly, Figure 17's first row shows the process of climbing down a single step of a 40 cm wide, 4 cm high and 20 cm deep straight staircase. The second row demonstrates the descent down the staircase's remaining four steps.

These figures show still frames of a video sequence where our robot successfully walks up or down a wooden bar and a straight staircase climbs up or down.

Figure 12: NAO steps over a 3.5 cm height and 2 cm depth wooden obstacle using planned whole body movement

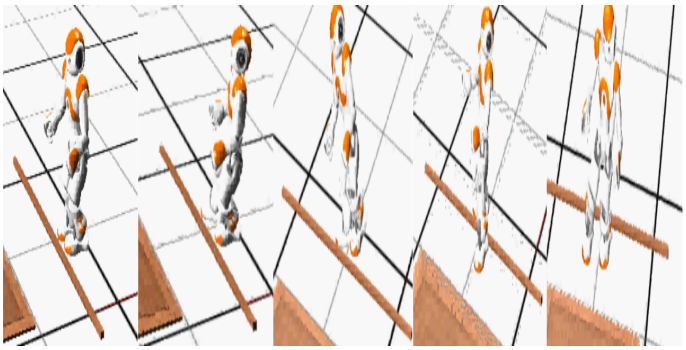


Figure 13: Simulated NAO walking over a 2 cm depth and 3.5 cm height wooden obstacle using expected whole body movement

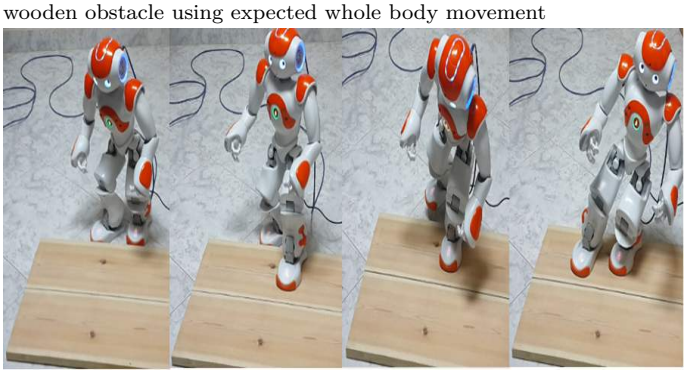


Figure 14: NAO step 2 cm high and 40 cm deep on a wooden obstacle using the expected whole body movement

Figure 15: NAO descending from a 2 cm height and 40 cm depth wooden obstacle using the expected whole body movement

Figure 16: NAO ascending 4 cm high and 20 cm deep stairs using the expected entire body movement

Figure 17: NAO descending 4 cm high and 20 cm deep stairs using the expected whole body movement

The algorithm applied for all motions is the same. The only exception is the height of the leg of NAO and the position of the swing foot, as in the case of walking up or down stairs and walking on or down obstacles, the swing foot is positioned closer to the standing foot. The justification for reducing the angle of positioning of the foot is the time and the opportunity to comfortably carry out more actions.

Our work performs a whole scenario starting with walking in search of a goal, including avoiding obstacles by climbing down straight stairs, climbing up stairs, as well as climbing over, on / down obstacles, until the goal is reached.

Table 2: Comparison between the proposed approach and other approaches

Motion Design	Foot Placement	Average Time	Obstacle Avoided	Vision
Maier et al. [3]	90°	2 min.	small bar, large bar	Depth Camera
Gouda et al. [24]	60°	43 sec.	Long bar, Wide bar	Monocular Camera
Designed Motion	30°	29 sec.	Long bar, Wide bar, straight staircase	Monocular Camera

The discrepancy between the suggested approach and the [3] approach is shown in the table 2. Declining foot displacement angle reduces the time as well as the robot's ability to perform more safely actions. The angle of the displacement of the foot can no longer be reduced as after 30°, since the robot could not perform actions.

The robot steps over it when a long bar is recognized and transfers his leg to the ground after the obstacle. Whereas the robot steps on/down the bar or stairs in the case of a broad or staircase and lifts his leg on or down it. The period for all motions to be performed is quite close. It takes the robot 30 seconds to execute step over movement, 29 seconds to execute step on / down movement, and 28 seconds to ascend one step up or down.

We perform a systematic evaluation of our approach to correctly walking over/onto / down an obstacle and climbing up stairs. Only on-board sensors are used to determine the success rate of these actions. In ten real world runs on our straight staircase consisting of four steps, approximately 96% of the straight staircase was successfully climbed up / down by the robot. Only one of the climbing steps contributed to the robot's crash.

The robot also successfully moved a total of eight subsequent times over/onto/down the wooden bar. The joints are then heated for an extended period of time by applying a stress on them. Joint overheating changes the parameters of the joints, especially rigidity, which affects the robot's balance; thus movements can no longer be performed effectively and the robot does not overcome the obstacle and sometimes falls. Through reducing the time spent in critical positions or setting stiffness to 0 after each operation, it is possible to reduce the heating in the joints.

Another problem is the execution time of the motion, as the robot must have enough time to reach balance after each action is performed in the motion or it will fall. The robot will not be able to finish the action it is doing in the event of the time being too short, so equilibrium will not be achieved and the robot will crash. However, if the execution time is too long to allow the robot to finish the preforming operation, its joints may easily get hot and may not be able to keep their balance in each position for a long time, so it may also fall.

These results show that our method makes it possible for the humanoid NAO H25 to climb up or down reliably four subsequent steps of straight staircases that are not visually detectable easily. In addition, by stepping over/onto/down them efficiently, the robot can avoid colliding with ground obstacles.

4 Conclusion

Throughout this research, we presented an innovative approach that allows a humanoid robot, particularly NAO H25, to design and implement entire body stability set of actions such as stepping over and overcoming obstacles, as well as ascending or descending a staircase using only on-board sensing. NAO is capable of walking on multiple floor surfaces such as carpets, tiles and wooden floors, it is capable of walking between these surfaces.

Large obstacles, however, can still cause him to fall because it assumes that the ground is more or less flat. The construction of complex dynamic movements can only be accomplished by using the own kinematics of the robot. Based on the recognized object obtained from the robot's database using the robot's own monocular camera (using NAO's vision recognition module), a sequence of actions to avoid the object is being executed.

One of NAO's most important challenges is to accurately identify the proper sequence of motions that help it to safely avoid obstacles without colliding with them. NAO is not able to use its standard motion controller to perform complex motions such as obstacle avoidance, two motion designs are investigated and presented to overcome this limitation.

Another difficulty is to assess the distance between the robot and the obstacle. To robots, distance measurement is

very important as the robot has to go to the exact position to perform additional tasks, such as playing with a ball, controlling the room, etc.

The robot's vision recognition module is partly reliable as it varies from half to up to twice the distance used for learning to distance between the robot and the target. The camera of the robot also has the limitation that it can not provide the distance between the robot and the obstacle; in order to overcome this limitation, a simple solution is implemented to allow the robot to learn images for the object on the floor in a place close to the robot's feet in the learning phase next to images for the floor itself. In deciding the distance between the robot and the obstacle, this solution proves to be very efficient.

This work is a major extension of the work discussed in [24, 1]. Shown in the simulation as well as actual world experiments with NAO H25 humanoid, in congested, multilevel environments with obstacles of different shapes and sizes, the robot can effectively execute whole-body movements.

References

[1] W. Gouda and W. Gomaa, "Complex motion planning for nao humanoid robot," in *11th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 2014.

[2] C. Graf, A. Härtl, T. Röfer, and T. Laue, "A robust closed-loop gait for the standard platform league humanoid," in *Proceedings of the Fourth Workshop on Humanoid Soccer Robots in conjunction with the*, pp. 30–37, 2009.

[3] D. Maier, C. Lutz, and M. Bennewitz, "Integrated perception, mapping, and footstep planning for humanoid navigation among 3d obstacles," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pp. 2658–2664, IEEE, 2013.

[4] F. Burget, A. Hornung, and M. Bennewitz, "Whole-body motion planning for manipulation of articulated objects," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pp. 1656–1662, IEEE, 2013.

[5] M. Mattamala, G. Olave, C. González, N. Hasbún, and J. Ruiz-del Solar, "The nao backpack: An open-hardware add-on for fast software development with the nao robot," in *Robot World Cup*, pp. 302–311, Springer, 2017.

[6] S. Shamsuddin, L. I. Ismail, H. Yussof, N. Ismarrubie Zahari, S. Bahari, H. Hashim, and A. Jaffar, "Humanoid robot nao: Review of control and motion exploration," in *Control System, Computing and Engineering (ICCSCE), 2011 IEEE International Conference on*, pp. 511–516, IEEE, 2011.

[7] S. Oßwald, A. Gorog, A. Hornung, and M. Bennewitz, "Autonomous climbing of spiral staircases with humanoids," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pp. 4844–4849, IEEE, 2011.

[8] A. Choudhury, H. Li, D. Greene, S. Perumalla, *et al.*, "Humanoid robot-application and influence," *Avishek Choudhury, Huiyan Li, Christopher M Greene, Sunanda Perumalla. Humanoid Robot-Application and Influence. Archives of Clinical and Biomedical Research*, vol. 2, no. 2018, pp. 198–227, 2018.

[9] A. Hornung, K. M. Wurm, and M. Bennewitz, "Humanoid robot localization in complex indoor environments," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp. 1690–1695, IEEE, 2010.

[10] W. Gouda, W. Gomaa, and T. Ogawa, "Vision based slam for humanoid robots: A survey," in *Electronics, Communications and Computers (JEC-ECC), 2013 Japan-Egypt International Conference on*, pp. 170–175, IEEE, 2013.

[11] G. Majgaard, "Humanoid robots in the classroom.," *IADIS International Journal on WWW/Internet*, vol. 13, no. 1, 2015.

[12] S. Kucuk and Z. Bingul, "Robot kinematics: forward and inverse kinematics," *Industrial Robotics: Theory, Modeling and Control*, pp. 117–148, 2006.

[13] M. Gienger, M. Toussaint, and C. Goerick, "Whole-body motion planning–building blocks for intelligent systems," in *Motion Planning for Humanoid Robots*, pp. 67–98, Springer, 2010.

[14] N. Kofinas, *Forward and inverse kinematics for the NAO humanoid robot*. PhD thesis, Diploma thesis, Technical University of Crete, Greece, 2012.

[15] M. Szumowski, M. S. Żurawska, and T. Zielińska, "Preview control applied for humanoid robot motion generation," *Archives of Control Sciences*, vol. 29, 2019.

[16] A. Gholami, M. Moradi, and M. Majidi, "A simulation platform design and kinematics analysis of mrl-hsl humanoid robot," 2019.

[17] M. Assad-Uz-Zaman, M. R. Islam, and M. H. Rahman, "Upper-extremity rehabilitation with nao robot," in *Proceedings of the 5th International Conference of Control, Dynamic Systems, and Robotics (CDSR'18)*, pp. 117–1, 2018.

[18] AldebaranRobotics, "Nao software 1.14.5 documentation @ONLINE," 2014.

[19] C. Bowen, G. Ye, and R. Alterovitz, "Asymptotically optimal motion planning for learned tasks using time-dependent cost maps," 2015.

[20] S. Shin and C. Kim, "Human-like motion generation and control for humanoid's dual arm object manipulation," 2014.

[21] K. Nishiwaki, S. Kagami, Y. Kuniyoshi, M. Inaba, and H. Inoue, "Toe joints that enhance bipedal and fullbody motion of humanoid robots," in *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, vol. 3, pp. 3105–3110, IEEE, 2002.

[22] J. E. Chestnutt, K. Nishiwaki, J. Kuffner, and S. Kagami, "An adaptive action model for legged navigation planning.," in *Humanoids*, pp. 196–202, 2007.

[23] W. Samakming and J. Srinonchat, "Development image processing technique for climbing stair of small humaniod robot," in *Computer Science and Information Technology, 2008. ICCSIT'08. International Conference on*, pp. 616–619, IEEE, 2008.

[24] W. Gouda and W. Gomaa, "Nao humanoid robot motion planning based on its own kinematics," in *Methods and Models in Automation and Robotics (MMAR), 2014 19th International Conference On*, pp. 288–293, IEEE, 2014.

[25] J. Guan and M.-H. Meng, "Study on distance measurement for nao humanoid robot," in *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*, pp. 283–286, IEEE, 2012.

[26] Cyberbotics, "Webots: the mobile robotics simulation software @ONLINE," 2014.

[27] L. D. Evjemo, "Telemanipulation of nao robot using sixense stem-joint control by external analytic ik-solver," Master's thesis, NTNU, 2016.

[28] RobotsLAB, "Nao evolution - v5," 2015.

[29] T. Deyle, "Aldebaran robotics announces nao educational partnership program," 2010.

[30] N. Kofinas, E. Orfanoudakis, and M. G. Lagoudakis, "Complete analytical inverse kinematics for nao," in *Autonomous Robot Systems (Robotica), 2013 13th International Conference on*, pp. 1–6, IEEE, 2013.

[31] E. Pot, J. Monceaux, R. Gelin, and B. Maisonnier, "Choregraphe: a graphical tool for humanoid robot programming," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pp. 46–51, IEEE, 2009.

[32] L. Ismail, S. Shamsuddin, H. Yussof, H. Hashim, S. Bahari, A. Jaafar, and I. Zahari, "Face detection technique of humanoid robot nao for application in robotic assistive therapy," in *Control System, Computing and Engineering (ICCSCE), 2011 IEEE International Conference on*, pp. 517–521, IEEE, 2011.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 1, 285-293 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Implementation of Paraconsistent Logic Based PI Controller for TA Converter

Tapas Kumar Mohapatra*, Asim Kumar Dey, Krushna Keshab Mohapatra

SOA University, Department of Electrical Engineering, 751030, India

ARTICLE INFO

Article history:
Received: 14 October, 2019
Accepted: 23 January, 2020
Online: 10 February, 2020

Keywords:
Photo Voltaic (PV)
Paraconsistent Logic (PL)
Positive output voltage
TA converter
Paraconsistent logic based PI (PLPI)

ABSTRACT

A novel control technique is implemented for TA converter. This converter is used for charging of electric vehicle Battery from a Photo Voltaic (PV) Module. TA converter is one type of non-isolated buck-boost converter. This converter provides positive output voltage. This article is an extension work that is originally presented in 2019 IEEE International Conference on Sustainable Energy Technologies (ICSET). A Comparative study of the performances of different converters has been described. The Paraconsistent Logic (PL) is used for a PI controller. The performance of new Paraconsistent Logic based PI (PLPI) controller is compared with the conventional PI controller. This control has the similar action as PI and the logical actions as structured in Paraconsistent Logic. Operational details of TA converter with the different performances has been described in this article. MATLAB is used to validate the performance by developing both state space and simulation model. The robustness of the controller is compared with traditional PI controller. The hardware of the TA converter with the new PLPI control technique is projected and also matched with simulation results.

1 Introduction

Because of extraordinary utilization of petroleum derivative and air contamination, the auto mobile sectors are compelled to consider electric vehicles (EV). Presently a days EV battery modelling and its charging is an extraordinary research region. It is notable, switch mode power supply (SMPS) is the core of power transformation innovation. EV battery charging is promising through simple and efficient DC-DC converter. Mostly buck-boost converter is preferred for the EV battery charging process. Buck and boost converters are simple and efficient but unable to provide high voltage gain. Multiple circuits have been projected in different articles such as the complex topology developed by LUO [1]–[3] provides a high voltage gain. It is costly, big in size and losses is also more. Interleaved converters [4]–[10] are able to provide high voltage gain with less voltage stress but their operational mode ,converter arrangement and control strategies are very complex. Hwu and Peng [10] also developed buck-boost topology which can provide voltage gain of 2D with positive output. Quadratic converters [11] – [14] can provide high voltage gain with less efficiency. KY converter [15] has high voltage gain but increases complexity with additional switches and also increases the overall cost.Though the traditional buck-boost converter introduce high efficiency with reduced cost but unable to provide higher and positive output voltage. In 1991 Cuk suggested a converter, [16] which is able to provide a voltage gain of $\frac{D^2}{(1-D)^2}$ but can only operate in buck mode due to the presence of clamping diode D_1 and D_2. The converter which provides low gain can provide high or low voltage output at very high or very low duty ratio, which is practically difficult to achieve. A new converter proposed in [17], having little voltage ripple, negligible radio frequency intervention, and one shared ground switch but complexity increases due to the 7^{th} circuit and also having different ground in its input and output terminals. In [18], a cascaded converter, combining two separated converters with both current source and sink, is functional for the thermoelectric generator. However, the voltage gain is also unnatural. Particularly, in order to attain high-voltage gain, the above converters must be functioning under extremely high or little duty cycle which is difficult to realize due to the practical limitations. Therefore, discovering new converter topology to overcome the negatives of the conventional ones for the requirements in industries is significant and appreciated.

TA converter [19] is proposed in the article which is one type non isolated positive output based buck-boost converter. It introduce a voltage gain of $\frac{D^2}{(1-D)^2}$. Though the voltage gain is similar to several other proposed converters but it disables few problems associated with others. This TA converter has several advantages over Shan and Faqiang converter [20] like:

*Tapas Kumar Mohapatra , Dept. of Electrical Engineering, SOA University & tapasmohapatra@soa.ac.in

- High voltage gain with positive output voltage.
- Less current stress in capacitor.
- Less voltage stress in inductor.
- Less source current for same load current in comparison to others.
- Can operate at CCM mode at low switching frequency.

The basic range of battery used for electric bike varies between 12 to 48 volt. In this article a battery of 12 volt, 7AH rating is used as a load, which is charged from a source varies between 6 to 30 volt. TA converter with PI controller is used to provide a constant voltage of 12 volt from a solar panel.

A controller/regulator is basically act as the brain of the system. Three types of conventional controller such as Proportional, Proportional and Integral (PI) and Proportional, Integral and Derivative (PID) have a wide scope of uses in SMPS application. With scientific advancement, controllers got computational assets about the 1980s, and turned out to be increasingly productive. The incorporation of micro controller chips for the implementation of controllers, in DC to DC converters, allowing the integration of algorithms with P, PI and PID activities, the expansion of structures with automatic tuning and massive simplicity in control parameter settings [21] - [24]. The primary goal of this work is to introduce the consequences of the PI controller using PL in the novel DC to DC boost converter and also compare the result with the conventional PI controller.

In section II of the article, the converter circuit diagram, operation and analysis is explained. Mathematical model and Simulink model is revealed in section III. In section IV, the Paraconsistent Logic based PI controller (PLPI) is explained. In section V hardware and simulation result of the converter is compared for PL-PI as well as the results compared with other converters. In the section VI the dynamic performance of the proposed controller is compared with conventional PI controller. Few comments and concluding remarks are given in section VII

2 Proposed TA converter Structure, Operation and analysis

2.1 Structure / Circuit Diagram

TA converter circuit diagram is shown in figure-1. The circuit elements of the converter are as follows: S_1 and S_2) - Power switches,(D_1 and D_2) – Diodes, (L_1 and L_1)- Inductors, (C_1 and C_1) - Capacitors. Time domain wave forms of the TA converter in CCM is shown in figure-3.

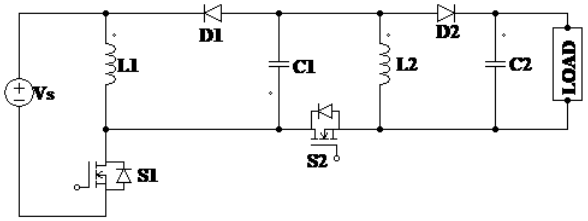


Figure 1: Circuit diagram of TA converter

2.2 Operating Principle and Analysis

In continuous conduction mode (CCM) the converter drives in two modes. First Mode -(0 < t < DT): In first mode S_1 and S_2 switches are simultaneously turned on and acts as short circuit, D_1 and D_2 diodes get reverse biased and act as open circuit. The direction of inductor currents are shown in figure - 2(a). C_2 provides the necessary output power to load. The following dynamic equations (1 - 8 will be obtained from figure - 2(a).

$$V_s - L_1\frac{di_{L1}}{dt} = 0 \quad (1)$$

$$\int_{I_{L1min}}^{I_{L1max}} di_{L1} = \int_0^{DT} \frac{V_s}{L_1} \quad (2)$$

$$\Delta I_{L1} = \frac{V_S}{L_1}DT \quad (3)$$

$$V_s - V_{L1} = 0 \quad (4)$$

$$V_{C1} - L_2\frac{di_{L2}}{dt} = 0 \quad (5)$$

$$\int_{I_{L2\min}}^{I_{L2\max}} di_{L2} = \int_0^{DT} \frac{V_{C1}}{L_2}dt \quad (6)$$

$$\Delta I_{L2} = \frac{V_{C1}}{L_2}DT \quad (7)$$

$$V_{C1} - V_{L1} = 0 \quad (8)$$

Second Mode - (0 < t < (1-D)T): In second mode S_1 and S_2 switches are simultaneously turned off and acts as open circuit, D_1 and D_2 diodes get forward biased by the inductors and act as short circuit. The direction of inductor currents are shown in Figure - 2(b). Inductor provides the necessary output power to load and to charge the capacitor C_2. The following dynamic equations (9 - 16) will be obtained from figure-2(b).

$$-V_{C1} - L_1\frac{di_{L1}}{dt} = 0 \quad (9)$$

$$\int_{I_{L1\max}}^{I_{L1\min}} di_{L1} = \int_0^{(1-D)T} \frac{-V_{C1}}{L_1}dt \quad (10)$$

$$\Delta I_{L1} = \frac{V_{C1}}{L_1}(1-D)T \quad (11)$$

$$-V_{C1} - V_{L1} = 0 \quad (12)$$

$$-V_{C2} - L_2\frac{di_{L2}}{dt} = 0 \quad (13)$$

$$\int_{I_{L2\max}}^{I_{L2\min}} di_{L2} = \int_0^{(1-D)T} \frac{-V_{C2}}{L_2}dt \quad (14)$$

$$\Delta I_{L2} = \frac{V_{C2}}{L_2}(1-D)T \quad (15)$$

$$-V_{C2} - V_{L2} = 0 \quad (16)$$

As we know the inductor voltage over a period is zero. We will obtain the following relations as given in equations (17-19). The inductor currents and capacitor voltages ripple are written in equation (20-23). Where f_{SW} is the switching frequency.

$$V_{C2} = V_0 = V_{C1}\frac{D}{1-D} \quad (17)$$

$$V_{C1} = V_S\frac{D}{1-D} \quad (18)$$

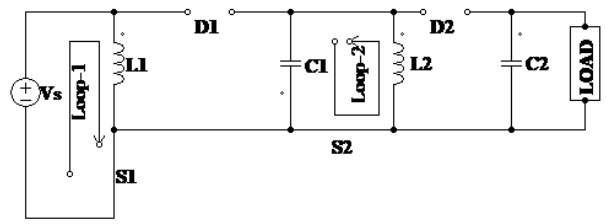
$$V_0 = V_S\frac{D^2}{(1-D)^2} \quad (19)$$

$$\Delta I_{L1} = \frac{V_S D}{L_1 f_{Sw}} \quad (20)$$

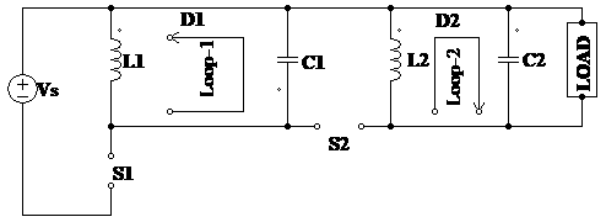
$$\Delta I_{L2} = \frac{V_S D^2}{L_2 f_{Sw}(1-D)} \quad (21)$$

$$\Delta V_{C1} = \frac{V_0 D}{RC_1 f_{Sw}(1-D)} \quad (22)$$

$$\Delta V_C 2 = \frac{V_0 D}{RC_2 f_{Sw}} \quad (23)$$



(a) First Mode



(b) Second Mode

Figure 2: Equivalent circuit diagram of two modes of TA converter

3 TA converter mathematical and simulink model

3.1 State Space Model

From the Mode-I and Mode-II the following dynamic equation will be obtained using which the state model can be obtained. The equations are given in (24-31).

$$\frac{di_{L1}}{dt} = \frac{V_S}{L_1} \quad (24) \qquad \frac{di_{L1}}{dt} = \frac{-V_{C1}}{L_1} \quad (28)$$

$$\frac{di_{L2}}{dt} = \frac{V_{C1}}{L_2} \quad (25) \qquad \frac{di_{L2}}{dt} = -\frac{V_{C2}}{L_2} \quad (29)$$

$$\frac{dV_{C2}}{dt} = -\frac{V_{C2}}{RC_2} \quad (26) \qquad \frac{dV_{C2}}{dt} = \frac{i_{L2}}{C_2} - \frac{V_{C2}}{RC_2} \quad (30)$$

$$\frac{dV_{C1}}{dt} = i_{L2} \quad (27) \qquad \frac{dV_{C1}}{dt} = -i_{L1} \quad (31)$$

$$\frac{di_{L1}}{dt} = D\frac{V_S}{L_1} - (1-D)\frac{V_{C1}}{L_1} \quad (32)$$

$$\frac{di_{L2}}{dt} = D\frac{V_{C1}}{L_2} - (1-D)\frac{V_{C2}}{L_2} \quad (33)$$

$$\frac{dV_{C2}}{dt} = -D\frac{V_{C2}}{RC_2} + (1-D)\frac{i_{L2}}{C_2} - (1-D)\frac{V_{C2}}{RC_2} \quad (34)$$

$$\frac{dV_{C1}}{dt} = Di_{L2} - (1-D)i_{L1} \quad (35)$$

The average model is given in equations (32-35).From the above equation it is clearly visible that the system is a fourth order one.

3.2 Simulation model and results

Figure-4 shows the Simulink model of the projected buck boost converter with PLPI controller. The circuit parameters that are used in the Simulink model are given in Table-II. Only one gate signal is essential for both the switches as both of them are operating simultaneously. A source of 18 volt is applied with output voltage of 40 volt in boost mode and 14 volt in buck mode. Input voltage is varied from 4volt to 18volt in boost mode and 18 volt to 100 volt in buck mode for the verification of robustness. Load resistance is also varied from 40Ω to 160Ω in boost and 5Ω to 50Ω in buck mode for the same verification. The load voltage remains as required with the above voltage and load resistance variations. In the figure -5 currents of each inductor and source, voltages of each capacitor, switch and diodes in boost mode are shown. Similarly in figure -6 the above waveforms in buck mode are displayed. It is clearly visible from figure - 5 and figure - 6 that i_{L1} never goes to negative value for both mode of operation.

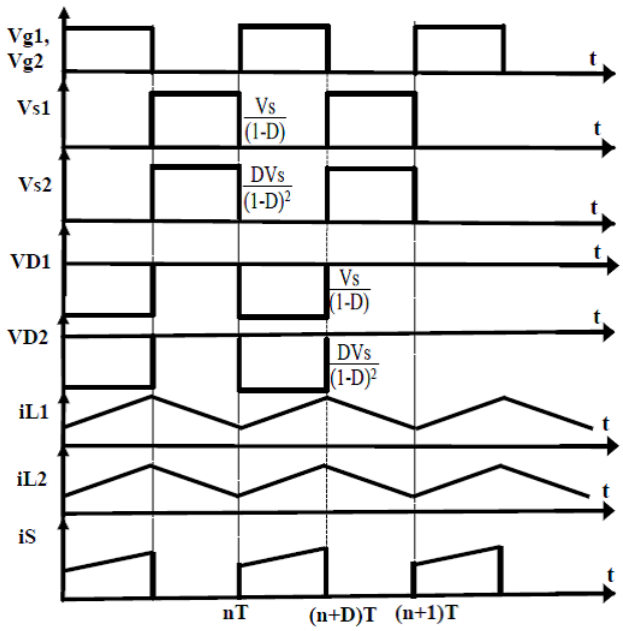


Figure 3: Different voltage and current wave forms

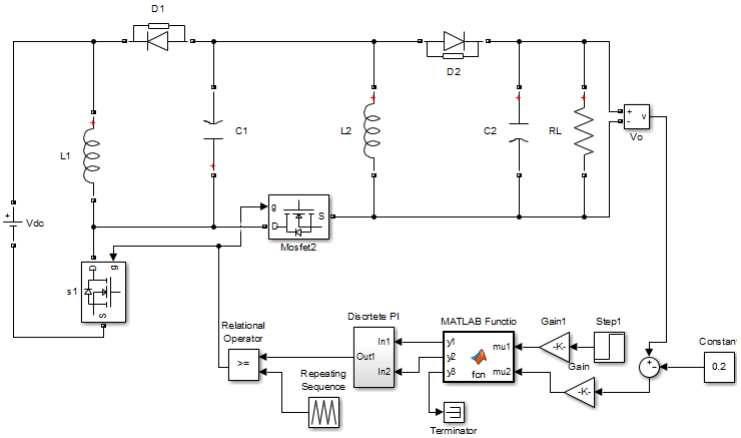


Figure 4: Matlab simulink model of proposed TA converter

4 Paraconsistent logic based PI controller

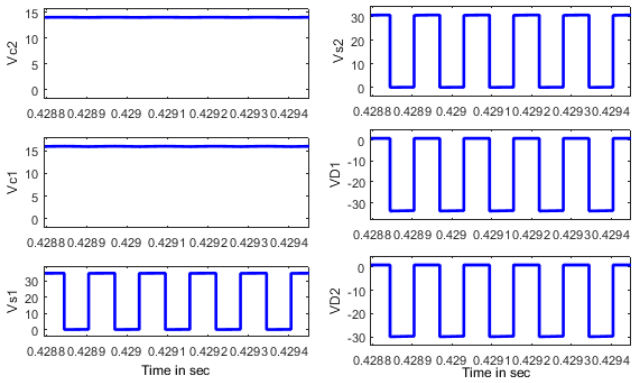
4.1 Logic of paraconsistent.

The conventional or Aristotelian logic underpins our present innovation is created on rigid binary laws and, along these lines, does not concede circumstances of redundancy, irregularities or those that are communicated by deficiency [25]. In conventional logic there is no logical inconsistency like something can't be both genuine and not valid in the meantime, when dealing with similar context. In any

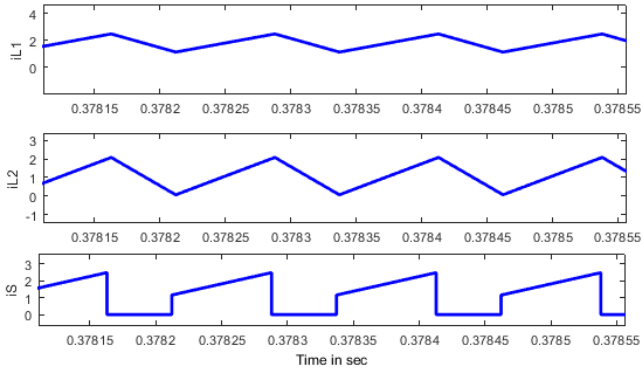
case, with the cut-off points of measures of control and eminence at present mandatory, progressively polished, make the philosophies of Conventional logic difficult to give effective control prototypes so improved for this mandate [25]. So as to act in circumstances where binary logic is difficult to be connected, different sorts of logic, called non-conventional, for example, Fuzzy, multivalued, paraconsistent, and so on, have as of late been made [25] - [26].

Non-conventional logics have the primary target is to restrict the binary principle of conventional logic and also catch diverse parts of casual conflicts. For instance, the principle of non-contradiction (PNC) in conventional logic explains that conflicting articulations can't both be valid in a similar sense in the meantime. Be that as it may, on account of signals as for a physical quantity beginning from two sensors, they might be conflicting, which gives conflicting data to make decision. In this way, control system will be progressively effective on the off chance that they are ready to act in circumstances where data might be conflicting.

A non-conventional Paraconsistent (PL) logic property states that it can tolerate the logical inconsistency in its establishments and is capable of managing conflicting signals. The PL has as its expansion the Paraconsistent Annotated Logic (PAL), which has a Hasse (Lattice FOUR) related with logical states existing at its vertices. Along these lines, sentences can be acquired where recommendations can be examined based on confirmations. In this PAL portrayal, the four extraordinary logic existing at the vertices of the PAL lattice are: Paracomplete ($\perp$), False (f), True (t) and Inconsistent ($\top$) [27] - [28].

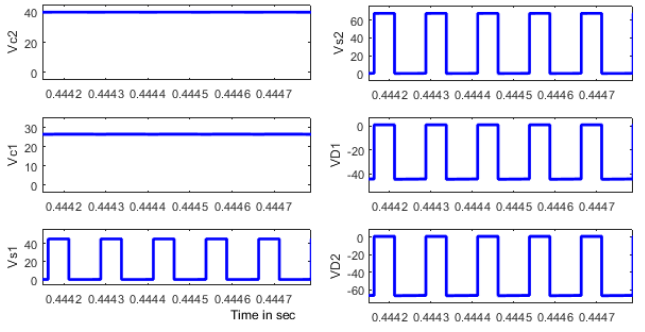


(a) $V_{c2}, V_{c1}, Vs1, V_{s2}, V_{D1}$ and V_{D2}

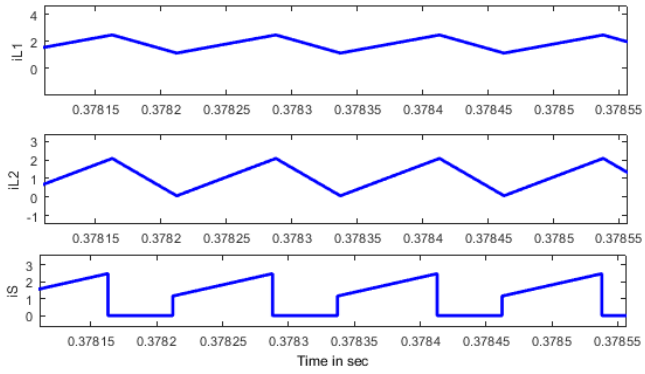


(b) $i_{L1}, i_{L2} and i_S$

Figure 6: Simulation result of TA converter in buck mode

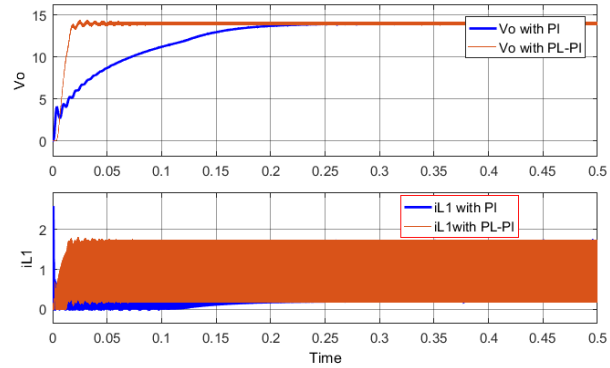


(a) $V_{c2}, V_{c1}, Vs1, V_{s2}, V_{D1}$ and V_{D2}

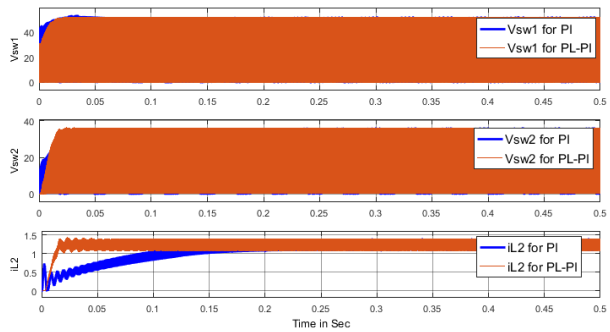


(b) $i_{L1}, i_{L2} and i_S$

Figure 5: Simulation result of TA converter in boost mode



(a) V_0 and i_{L1}



(b) $V_{sw1}, V_{sw2} and i_{L2}$

Figure 7: Comparison of results with PI and PL-PI

4.2 Logic with annotation of two values.

From a Lattice of four vertices it is conceivable to relate a sort of Paraconsistent Annotated Logic in which two evidential qualities related to a specific suggestion are considered. By translating the evidential values by methods for comment or standardized degrees of proof we can acquire conditions including the logical states that are characterized with degrees of proof from estimations. This sort of logic is called Paraconsistent Annotated Logic with explanation of two values (PAL2v) [27] - [28]. Figure-8(a) demonstrates the Lattice FOUR related at PAL2v.

In PAL2v an evidential nuclear equation of the shape P (μ , λ) can be viewed as an explanation for the recommendation p, where μ , $\lambda \in [0, 1]$ (genuine unit interim) [26] - [27]. Along these lines, the Evidence degree (μ) is an esteem that speaks to the proof ideal to the recommendation p, and the Evidence degree (λ) is an esteem that speaks to the proof adverse to the recommendation p. The relationship of a couple (μ ,λ) with a suggestion p implies that the level of ideal proof at p is μ , and the level of adverse proof at p is λ. As indicated by the comments in the related lattice appeared in figure-8(b)) [27]

- (1,1) → shows the presence of both, ideal and adverse proof totals, with a logical meaning of irregularity to the recommendation p.
- (1,0)→ shows the presence of complete ideal proof with adverse proof equal to zero, flagging an undertone of logical truth for the recommendation p.
- (0,0)→ shows the presence of ideal and adverse proof with the two values equal to zero, doling out a logical implication of uncertainty for the recommendation p.
- (0,1) → demonstrates the presence of ideal proof equal to zero and all out adverse proof, flagging a meaning of logical lie for the recommendation p.

The conditions of the PAL2v are acquired through a change where, at first, are viewed as characterised in a Unit Square in the Cartesian Plane (USCP) the level of ideal evidence μ in the y - axis, and the level of adverse evidence λ in the x - axis, as indicated by Figure-8(c)[26] - [28] A change that permits the degrees of proof of standardized qualities characterised in the x , y axis of the (USCP) to be situated on the X and Y axis of a four vertex PAL2v Lattice, given by equation - 36.

$$T(X, Y) = (xy, x + y - 1) \tag{36}$$

$$D_c = \mu - \lambda \tag{37}$$

$$D_{ct} = \mu + \lambda - 1 \tag{38}$$

$$\varepsilon\tau = (D_c D_{ct}) \tag{39}$$

$$\mu_{ctr} = (\mu + \lambda)/2 \tag{40}$$

$$d = \sqrt{(1 - |D_c|)^2 + (D_{ct})^2} \tag{41}$$

If $D_c > 0$

$$D_{CR} = 1 - \sqrt{(1 - |D_c|)^2 + (D_{ct})^2} \tag{42}$$

If $D_c < 0$

$$D_{CR} = \sqrt{(1 - |D_c|)^2 + (D_{ct})^2} - 1 \tag{43}$$

$$\mu_{ER} = (D_{CR} + 1)/2 \tag{44}$$

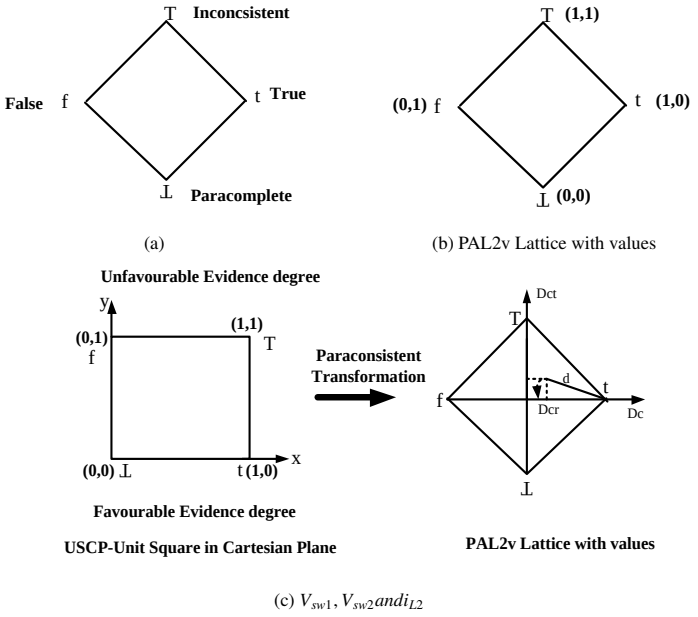


Figure 8: (a) Lattice associated with annotation of two values PAL2v. (b) PaL2v – Hasse finite Lattice with annotation. (c) Unit Square in the Cartesian Plane

Relating the parts of the change T(X , Y) concurring to the typical classification of PAL2v, where [26] - [28]: x = μ → Ideal Evidence degree, with $0 \leq \mu \leq 1$ and y = λ → level of Adverse Evidence, with $0 \leq \mu \leq 1$, where:

- The principal term acquired in the arranged match of the change condition is X = x-y = $\mu - \lambda$ → which will be known as the conviction degree D_c [27] - [29]. Accordingly, the conviction degree is acquired by equation - 37.
- The second term got in the arranged match of the change condition is: Y = x+y-1 = $\mu + \lambda$ - 1 → which will be known as the Contradiction Degree-Dct [27] - [29]. In this manner, the level of Contradiction is acquired by equation - 38.

The conditions of PAL2v permit the paraconsistent logical states to be found inside the lattice, the interpolation point is given by equation - 39. One can in this way settle on choices dependent on the vicinity of the logic state $\varepsilon\tau$ to the extraordinary logic states True (t) or False (f), found at the vertices of the PAL2v lattice. The Normalized Degree of Contradiction - μctr ,where a variety somewhere in the range of 0 and 1 is gotten, can be determined through condition [27] - [29] given in equation - 40. The Real Certainty degree D_{CR} is acquired by deciding the separation d in the PAL2v lattice as given in equation - 41. The (D_{CR}) values are determined by

the conditions appeared underneath [26] - [28] is given in equation - 42 and equation - 43 also linked by equation equation - 44.

The conditions gotten by the understandings made in the lattice related to PAL2v permit the making of algorithms utilized in examination and logical treatment of data signals. The Paraconsistent Analysis Node (PAN) algorithm can be utilized in a few fields of learning where fragmented and opposing data is dealt with fittingly through the PAL2v conditions. In this work, the PAN will be utilized to assemble a system of signals examination that embodies the control factors in a new converter.

4.3 Paraconsistent Algorithm

The explanation of a common PAN algorithm is specified under [26] - [28] State: P_1=input 1; P_2=input 2; O_1=outpu 1; O_2=output 2; O_3=output 3;

1 $\mu_1 = P_1$ (Ideal Evidence Degree), where $0 < \mu_1$. $\mu_2 = P_2$ (Ideal Evidence Degree), where $0 < \mu_2$.
$\lambda = 1 - \mu_2$ (Adverse Evidence Degree), where $0 < \lambda < 1$.

2 Evaluate the conviction degree $D_c = \mu 1 - \lambda$

3 Evaluate the degree of conviction $D_{ct} = \mu 1 + \lambda - 1$

4 Evaluate Normalized Degree of Contradiction. $\mu_{ctr} = (\mu 1 + \lambda)/2$

5 Evaluate the separation d in the PAL2v lattice. $d = \sqrt{(1 - |D_c|)^2 + (D_{ct})^2}$

6 Decide the output signals. If d > 1, then keep O_1=0.5; O_2= μ_{ctr} and $O_3 = D_{ct}$. Consider meaning less and go to exit. Else move to the next step

7 Decide actual conviction degree D_C. If $D_C > 0$ evaluate $D_{CR} = 1 - \sqrt{(1 - |D_c|)^2 + (D_{ct})^2}$
If $D_C < 0$ evaluate $D_{CR} = \sqrt{(1 - |D_c|)^2 + (D_{ct})^2} - 1$

8 Determine Resultant Real Evidence degree.
$\mu_{ER} = (D_{CR} + 1)/2$

9 The Real outputs are O_1=μ_{ER} , $O_2 = \mu_{ctr}$ and O_3= D_{ct} .

10 End.
In this article the Evidence degree was created as follows:
$0<\mu_2<1$ are proportional to $0.2V<P_2<14V$ and
$0<\mu_1<1$ are proportional to $0V<P_1<14V$.
$14V \rightarrow \mu_1 = 1$; with For $P_1 = 0V \rightarrow \mu_1 = 0$; and $P_1 = arelationof \mu_1 = P_1/14$.
$14V \rightarrow \mu_2 = 1$; with For $P_2 = 2V \rightarrow \mu_2 = 0$; and $P_2 = arelationof \mu_2 = (P_1 - 0.2)/14 and \lambda = 1 - \mu_2$.

5 Comparisons of Different Converters and the Control Technique

Table-I shows the comparisons of complexity, voltage gain, voltage and current stress, components of different converters such as traditional [30], Shan and Faqiang [20] and projected TA converter. Figure-9 curve shows the variation of voltage gain with the duty cycle. From observation it is clearly visible that TA converter provides a positive average inductor current but average inductor current under buck mode in Shan and Faqiang converter is negative. The peak inductor voltage V_{L2peak} and peak capacitor current I_{Cpeak} is less in TA converter. Figure-7 shows the comparison results of output voltage, inductor currents, switch voltages and source current between conventional PI and Paraconsistent logic based PI. From the figure it is clearly visible that the dynamic response of output voltage, inductor currents are faster in PL-PI than PI. It is also observed that for the same P and I gain the output voltage in PL-PI is obtained with less value of filter capacitor. The steady state values and time are same in both the controllers.

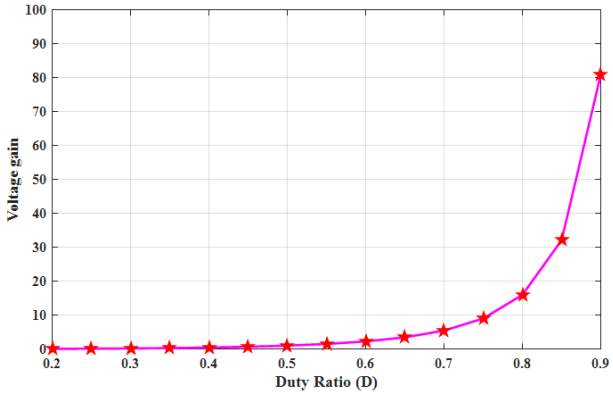


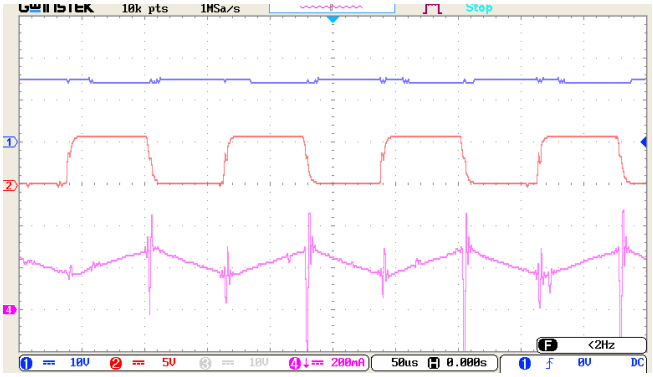
Figure 9: Voltage gain curve with duty ratio.

6 Hardware Parameters and Results

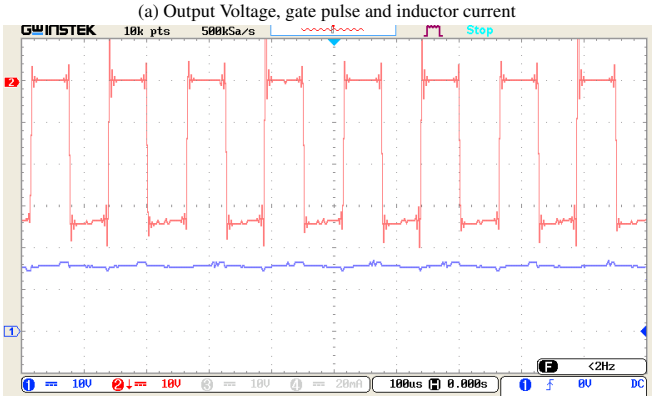
The parameters chosen for hardware implementation are provided in Table-II. PLPI controller is implemented for obtaining a fixed output voltage. The DSP (F28379D) Board is used for the designing of the proposed PLPI and traditional PI controller. The program is written in code composer studio and also the output file is burned on the DSP board using the same. Primarily a resistive load is taken to test the controller then the battery of rating 12volt 7Ah (Lead acid type) is taken as load. 1.4Amp of peak charging current is drawn by the battery. Figure-11 and Figure-10 shows the results in boost and buck mode respectively. Figure-12shows the photo of the complete hardware set up. Figure-13 which shows the efficiency curve explains clearly that efficiency in boost mode is higher than buck mode. The different results for battery load is shown in Figure-14.

7 Conclusion

A positive output based transformer less buck–boost (TA) converter is projected in this article. This converter overcomes several draw backs of the old-fashioned and Shan and Faqiang converter. The steady state operation principle, mathematical model, and performance comparisons with other converters are enlightened. The analysis is done using Matlab simulation and the verification of the same through experiment is done. This projected converter retains several merits in both boost and buck mode. This converter provides



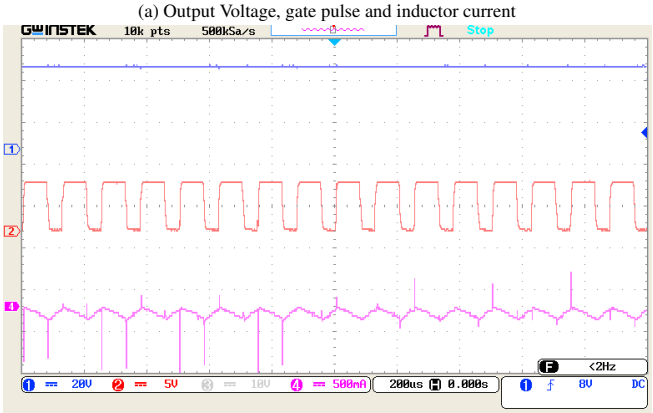
(a) Output Voltage, gate pulse and inductor current



(b) Capacitor and Diode Voltage

Figure 10: Buck mode results.

(a) Output Voltage, gate pulse and inductor current



(b) Capacitor and Diode Voltage

Figure 11: Boost mode results.

Figure 12: Picture of the hardware setup of the proposed TA Converter with DSP and PC interface.

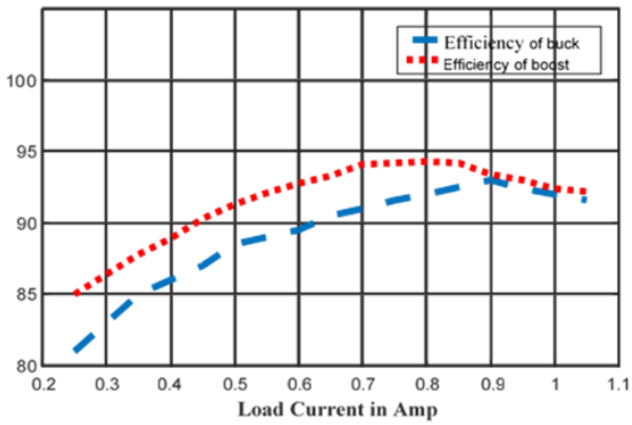


Figure 13: Efficiency Versus Load current for both Buck and Boost mode.

Figure 14: Output Voltage, gate pulse and inductor current for a Battery Load.

positive current through inductor and also provides positive output voltage. The converter control strategy is easy to implement. So, the planned converter is suitable for EV battery charging and industrial applications where high voltage gain is required. This article has also projected paraconsistent logic based PI controller. A comparison result is also shown in the figures. From the simulation and hardware result it is proved that the new control technique provides a better dynamic performance as well as provides better result with low value of capacitors.

References

[1] F. L. Luo and H. Ye, "Positive output cascade boost converters," Proc. Inst.Elect. Eng. Elect. Power Appl., vol. 151, no. 5, pp. 590–606, Sep 2004.

[2] Y. T. Chen,W. C. Lin, and R. H. Liang, "An interleaved high step-up DC–DC converter with double boost paths," Int. J. Circ. Theor. Appl., vol. 43, no. 8, pp. 967–983, Aug. 2015.

[3] L.W. Zhou, B. X. Zhu, Q. M. Luo, and S. Chen, "Interleaved non-isolated high step-up DC/DC converter based on the diode–capacitor multiplier,"IET Power Electron., vol. 7, no. 2, pp. 390–397, Feb. 2014.

[4] C. T. Pan, C. F. Chuang, and C. C. Chu, "A novel transformer less adaptable voltage quadrupler DC converter with low switch voltage stress," IEEE Trans. Power Electron., vol. 29, no. 9, pp. 4787–4796, Sep. 2014.

[5] K. I. Hwu, Member, IEEE, and Y. T. Yau, Student Member, IEEE" Two Types of KY Buck–Boost Converters" IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 56, NO. 8, AUGUST 2009.

[6] D. Maksimovic and S. Cuk, "Switching converters with wide DC conversion range," IEEE Trans. Power Electron., vol. 6, no. 1, pp. 151–157, Jan.1991.

[7] Pan CT, Lai CM. A high-efficiency high step-up converter with low switch voltage stress for fuel-cell system applications. IEEE Transactions on Industrial Electronics. 2010 Jun;57(6):1998-2006.

[8] Luo FL. Six self-lift DC-DC converters, voltage lift technique. IEEE Transactions on Industrial Electronics. 2001 Dec;48(6):1268-72.

[9] Ye YM, Cheng KW. Quadratic boost converter with low buffer capacitor stress. IET Power Electronics. 2014 May;7(5):1162-70.

[10] K. I. Hwu and T. J. Peng, "A novel buck–boost converter combining KY and buck converters," IEEE Trans. Power Electron., vol. 27, no. 5, pp. 2236– 2241, May 2012.

[11] Morales-Saldana JA, Loera-Palomo R, Palacios-Hernández E, Gonzalez-Martinez JL. Modelling and control of a DC-DC quadratic boost converter with R 2 P 2. IET Power Electronics. 2014 Jan;7(1):11-22.

[12] Hua CC, Chiang HC, Chuang CW. New boost converter based on Sheppard-Taylor topology. IET Power Electronics. 2014 Jan;7(1):167-76.

[13] Kim RY, Lai JS. Aggregated modeling and control of a boost-buck cascade converter for maximum power point tracking of a thermoelectric generator. InApplied Power Electronics Conference and Exposition, 2008. APEC 2008. Twenty-Third Annual IEEE 2008 Feb 24 (pp. 1754-1760). IEEE.

[14] Li, Wuhua, Xiaodong Lv, Yan Deng, Jun Liu, and Xiangning He. "A review of non-isolated high step-up DC/DC converters in renewable energy applications." In Applied Power Electronics Conference and Exposition, 2009. APEC 2009. Twenty-Fourth Annual IEEE, pp. 364-369. IEEE, 2009.

[15] Yang LS, Liang TJ, Chen JF. Transformer less DC–DC converters with high step-up voltage gain. IEEE Transactions on Industrial Electronics. 2009 Aug;56(8):3144-52.

[16] Miao Z, Fang LL. Enhanced self-lift Cuk converter for negative-to-positive voltage conversion. IEEE transactions on power electronics. 2010 Sep;25(9):2227-33.

[17] Hua CC, Chiang HC, Chuang CW. New boost converter based on Sheppard-Taylor topology. IET Power Electronics. 2014 Jan;7(1):167-76.

[18] Ajami A, Ardi H, Farakhor A. Design, analysis and implementation of a buck–boost DC/DC converter. IET Power Electronics. 2014 Dec 1;7(12):2902-13.

[19] Mohapatra, Tapas Ku, Asim Ku Dey, and K. K. Mohapatra. "A Novel Nonisolated Positive Output TA Converter." 2019 IEEE International Conference on Sustainable Energy Technologies (ICSET). IEEE.

[20] Shan Miao, Faqiang Wang and Xikui Ma, "A New Transformer less Buck–Boost Converter with Positive Output Voltage", IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 63, NO. 5, MAY 2016.

[21] Hwu KI, Yau YT. KY converter and its derivatives. IEEE Transactions on Power Electronics. 2009 Jan;24(1):128-37.

[22] Astrom KJ. T. haglund,"Auto Tuning of PID controllers," 1st ed. Research and triangle park, NC: instream. Soc. Amer. 1988.

[23] Zhao, D., Chai, T., and Wang H. Fu, J. (2014). Hybrid intelligent control for regrinding process in hematite beneficiation. Control Engineering Practice,22,217–230.

[24] Ogata, K. (2014). Modern control engineering (fifth ed.). Prentice Hall Inc, ISBN: 0-13-615673-8.

[25] Da Costa NC. On the theory of inconsistent formal systems. Notre dame journal of formal logic. 1974;15(4):497-510.

[26] Da Costa, N. C. A., and de Ronde, C. (2013). Found Phys, 43, 845. http://dx.doi.org/10.1007/ s10701-013-9721-9.

[27] Coelho MS, da Silva Filho JI, Côrtes HM, de Carvalho Jr A, Blos MF, Mario MC, Rocco A. Hybrid PI controller constructed with paraconsistent annotated logic. Control Engineering Practice. 2019 Mar 1;84:112-24.

[28] Baptista R, Da Silva Filho JI, Morilla JC. Proposal of Automatic Control three-way Valve through algorithms of Paraconsistent Annotated logic. Unisanta Science and Technology. 2013 Dec 14;2(2):69-73.

[29] Da Silva Filho JI. Algorithms based on paraconsistent annotated logic for applications in Expert systems. Expert System Software: Engineering, Advantages and Applications, Nova Science Publishers, Hauppauge. 2011:1-40.

[30] Y. He and F. L. Luo, "Analysis of Luo converters with voltage-lift circuit," Proc. Inst. Elect. Eng. Elect. Power Appl., vol. 152, no. 5, pp. 1239–1252, Sep. 2005.

Table 1: Comparison of different buck-boost converters

Parameters	Conventional Buck-Boost Converter [30]	Shan and Faqiang Converter [20]	TA converter [19]
Switches	1	2	2
Diodes	1	2	2
Capacitors	1	2	2
Inductors	1	2	2
Voltage gain	$\frac{D}{(1-D)}$	$(\frac{D}{(1-D)})^2$	$(\frac{D}{(1-D)})^2$
(Voltage stress on Switch-1)/V_S	$\frac{D}{(1-D)}$	$\frac{D}{(1-D)}$	$\frac{D}{(1-D)}$
(Voltage stress on Switch-2)/V_S		$(\frac{D}{(1-D)})^2$	$(\frac{D}{(1-D)})^2$
Nature of average Inductor current	Positive	Negative	Positive
complexity	2^{nd} order	4^{th} order	4^{th} order
Peak Voltage of L_2	V_C	V_{C1}+V_{C2}	V_{C2} or V_{C1}
Peak current through C_1	I_{Lmax}	I_{L1max} + I_{L2max}	I_{L1max} or I_{2Lmax}

Table 2: Different parameters of TA [19] converter

Parameters	Values	Parameters	Values	Parameters	Values
Inductor L_1	1 mH	V_S	18 Volt	Diodes	1N
Inductor L_2	3.5 mH	Proportional gain	0.0145	Load Resistance for Buck	5-50ω
Capacitor C_1	1μF (PL-PI) and 100μF (PI)	Integral gain	39.346	Load resistance for Boost	40 - 160 ω
Capacitor C_2	1μF (PL-PI) and 220-1000μF (PI)	Switches	MOFET IRF	Battery	Lead Acid 12V, 7AH

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 1, 294-301 (2020)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Mitigation Methods of Voltage Disturbances and Harmonic Currents of PWM Two-Level Inverter Supplying The 6-Pulse Diode Rectifier

Jerzy Ryszard Szymanski*, Marta Zurek-Mortka

Kazimierz Pulaski University of Technology and Humanities in Radom, Faculty of Transport, Electrical Engineering and Computer Science, Radom, Poland

ARTICLE INFO

Article history:
Received: 14 October, 2019
Accepted: 23 January, 2020
Online: 10 February, 2020

Keywords:
Common-mode voltage
Current harmonics
Differential-mode voltage
PWM modulation
Three-phase voltage inverter
Two-level PWM inverter

ABSTRACT

In fast charging station for Battery Electric Vehicles (BEVs) the three-phase PWM (Pulse Width Modulation) inverters generate parasitic Differential-Mode Voltages (DMV) and Common-Mode Voltages (CMV). Parasitic voltages are a side effect of using the width modulation to shape the phase-to-phase inverter's voltage. In this article the authors present a mathematical description of the differential-mode (DM) and common-mode (CM) voltages and carry out their spectral analysis. Based on the spectral harmonics analysis, the authors present a method for filtration of harmonics of DM and CM voltages aimed at limiting the capacitance parasitic currents: due to DM voltage – phase-to-phase parasitic current and CM voltage – ground parasitic currents. A resonant filter was used to eliminate harmonics of the input rectifier of the DC/DC converter supplied from a voltage inverter. As a result of tests, the THD_i harmonic content factor was significantly reduced and ground leakage current was eliminated from the PE electric shock protection system.

1 Introduction

The DC/DC converter rectifiers for charging electrochemical energy stores used in charging stations of Battery Electric Vehicles (BEV) are non-sinusoidal receivers and generate current harmonics. When PWM inverters are the source of voltage supplying local AC networks, it is necessary to eliminate high-frequency differential harmonics of differential-mode DM voltage and common-mode CM voltage, which are disturbances resulting from the use of PWM modulation. It is also necessary to limit the harmonic content of low rows in the phase current, as they increase losses in the network. In the case of rectifier power supply from the transformer, the phase and thus phase-to-phase voltages are deformed, transformer voltage deformations, determined by the THD_u harmonic content ratio, must not exceed of the permissible values [1],[2].

Scientific publications present solutions using transformers to galvanically separate the inverter, which converts the energy of the photovoltaic source from the input rectifier of the AC/DC/DC converter of the EV charger [3]. An interesting solution seems to be the use of a three-winding Yyd transformer in the 12-pulse diode rectifier of the AC/DC/DC charger [4]. One of the authors' research goals is to demonstrate the possibility of obtaining a level of harmonic content in the inverter current of a value close to when using a 12-pulse rectifier with a Yyd transformer.

This paper consist of theoretical part where the PWM inverter and the 6-pulse diode rectifier are analyses. It also shows the concept of powering from renewable energy sources dedicated for fast charging stations. Next part describes the solution of limiting disturbances of voltage inverters in hybrid system.Thanks to use of novel configuration of filters against voltage disturbances it is possible to mitigate the influence of rectifier on power supply system in charging station. The last part shows the results of simulations of rectifier's harmonics compensator and laboratory experiments to determine the conditions for cooperation of PWM converter with DC/DC converter using the CM and DM filters.

This article is an extension of the issues presented in the reference [5] and also is a continuation of the research presented there.

2 Hybrid Powering of Power Electronic Converters of EV Charging Station

In order to assure of high power and using of "clean" electricity, the electric vehicles charging station should be powered from hybrid system with an energy storage. Figure 1 shows an exemplary implementation of a hybrid power supply system for electric vehicles charging station with using the renewable energy sources.

In a hybrid power system a few energy sources is used. The loads are supplying from local AC or DC voltage supply network.

*Corresponding Author: Jerzy Ryszard Szymanski, phone: +48 601 292 535, e-mail: j.szymanski@uthrad.pl

In the hybrid network shown in Figure 1, the local DC voltage line is connected to renewable energy sources (RES) (e.g. wind farms or solar power plants) and energy storages, like batteries, fuel cells or pumped storages. The AC line is powering also DC receivers which are DC fast charging stations for electric vehicles. Solutions are known in which electric vehicles are connected charging station supplies energy to the power grid, here two-way DC/DC converters are used as chargers for electric vehicles EV. The energy of electrochemical batteries can also be sent to an industrial power supply network. At the same time, a power system that is synchronized with the local hybrid power network can complement local sources of "clean" energy from periods of reduced efficiency (e.g. on cloudy days for solar sources). In periods of increased efficiency of local renewable energy sources, excess energy can improve the efficiency of the power system. In such systems, bidirectional energy converters must be used [5], [6].

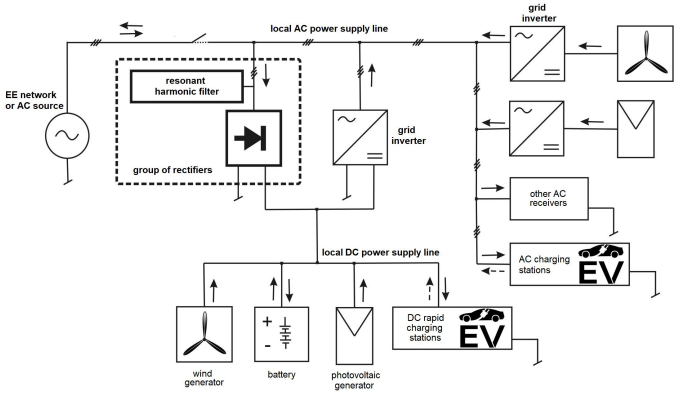


Figure 1: Local hybrid power supply system for EV charging stations with renewable energy sources (wind and PV)

Figure 2 shows the EV fast battery charging station, which is powered from a two-level voltage inverter or from the power grid. The passive parallel harmonic current filter is connected to the AC line with increased current consumption by the diode rectifier. This allows you to control the amount of capacitive reactive power supplied to the AC network by this filter. The passive filter used limits the level of harmonics in the THD_i range from 5% to 10% [5]. Powering the EV fast charging station from the low-voltage industrial network leads to distortion of phase and phase-to-phase voltages in this network. Voltage distortions increase due to the power of the battery charging station, in particular at similar capacities of MV / LV transformers in the power system. Such voltage distortions cause disturbances causing incorrect operation of other receivers connected to the power supply network, e.g. contactors, PLC controllers and other [7]. Limiting the current harmonic content is necessary today, because the number of diode rectifiers and their power is constantly increasing.

For the purposes of this article, it was assumed that the MV/LV transformer of the power system supplies, apart from the fast EV battery charging station, also other electric receivers. In places where travelers service on expressways and highways, where fast EV battery charging stations are located, transformers usually power other receivers, such as pumps, lighting, air conditioners. It was assumed that the fast charging station does not have its own dedicated MV/LV transformer, which is why the use of effective limiting harmonics of phase currents is very important.

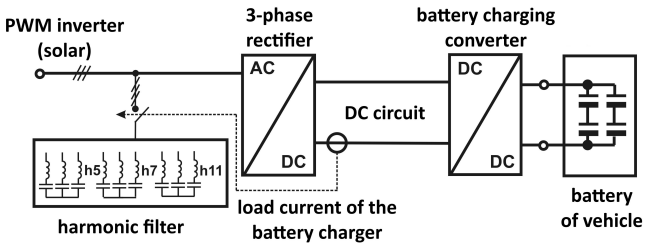


Figure 2: Block diagram of an EV battery charger with a passive harmonic current filter to provide sinusoidal current in the local power supply network

3 Mathematical Description of a Linear Three-Phase Inverter Model

The transformation of DC voltage into single-phase and three-phase AC voltage, with the use of the power-electronic PWM inverter, has side effects such as the CM voltages [8], [9]. In this chapter, the mathematical relations describing the voltage of the common mode disturbances were derived bases on the description of the three-phase linear model of the PWM inverter. The three-phase two-level voltage inverter is shown with a symmetrical load in Figure 3. It consists of three identical branches of half-bridges in each phase and this three-phase system is symmetrical for individual harmonics. In modeling and analyzing the inverter, a three-phase notation is used [10], [11].

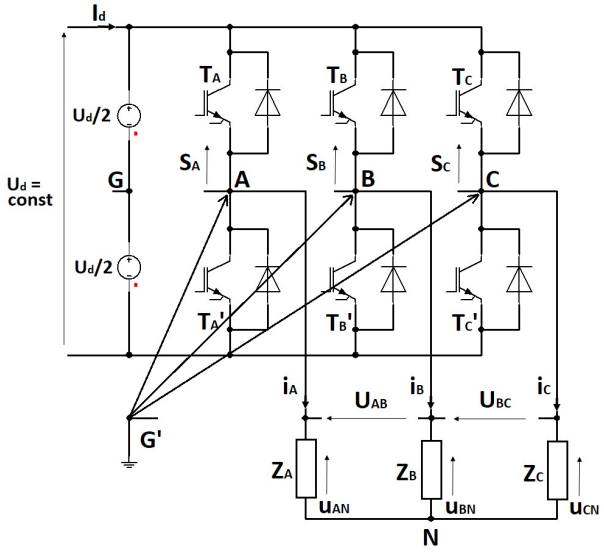


Figure 3: Basic scheme of three-phase two-level inverter with symmetrical load

On the basis of the three-phase inverter description, a mathematical relationship is determined, which describe the CM voltage of the inverter with the grounded negative DC voltage bus and without

galvanic ground of the power-electronic elements (such as an industrial inverters and auxiliary railway equipment inverters) [10], [12].

The output phase-to-phase voltages of the inverter are described as follow 1:

$$\begin{aligned} u_{AB} &= u_{AG} + u_{BG} \\ u_{BC} &= u_{BG} + u_{CG} \\ u_{CA} &= u_{CG} + u_{AG} \end{aligned} \tag{1}$$

where: u_{AG}, u_{BG}, u_{CG} are instantaneous voltages, which satisfy the condition 2:

$$u_{AB} + u_{BC} + u_{CA} = 0 \tag{2}$$

If the load neutral point N is not connected to the reference point G on the DC voltage side (Figure 3), then the phase voltages can be expressed as 3:

$$\begin{aligned} u_{AN} &= \tfrac{1}{3}(2u_{AG} - u_{BG} - u_{CG}) \\ u_{BN} &= \tfrac{1}{3}(-u_{AG} + 2u_{BG} - u_{CG}) \\ u_{CN} &= \tfrac{1}{3}(-u_{AG} - u_{BG} + 2u_{CG}) \end{aligned} \tag{3}$$

By describing the switching states of inverter S_A, S_B, S_C with 0 and 1 (e.g. S_A=1, indicates the state in which the transistor in phase A is switched on, and S_A=0, when the lower transistor in phase A is switched off), the output phase-to-phase voltages can be presented as follows 4:

$$\begin{aligned} u_{AB} &= U_d(S_A - S_B) \\ u_{BC} &= U_d(S_B - S_C) \\ u_{CA} &= U_d(S_C - S_A) \end{aligned} \tag{4}$$

Similarly for a symmetrical load the phase voltages are expressed by 5:

$$\begin{aligned} u_{AN} &= \tfrac{U_d}{3}(2S_A - S_B - S_C) \\ u_{BN} &= \tfrac{U_d}{3}(-S_A + 2S_B - S_C) \\ u_{CN} &= \tfrac{U_d}{3}(-S_A - S_B + 2S_C) \end{aligned} \tag{5}$$

If the G point is grounded in the Figure 3, then the u_{NG} voltage is determined by CM voltage of inverter 6:

$$\begin{aligned} u_{AG} &= u_{AN} + u_{NG} \\ u_{BG} &= u_{BN} + u_{NG} \\ u_{CG} &= u_{CN} + u_{NG} \end{aligned} \tag{6}$$

The sum of all voltages from equations 5 is 6:

$$u_{AG} + u_{BG} + u_{CG} = u_{AN} + u_{BN} + u_{CN} + 3u_{NG} \tag{7}$$

Since $u_{AG} + u_{BG} + u_{CG} = 0$, the CM voltage u_{NG} of inverter is determined by 8:

$$u_{NM} = \frac{u_{AN} + u_{BN} + u_{CN}}{3} \tag{8}$$

The relation 8 indicates that the CM voltage of inverter is equal to the voltage with respect to another ground point, e.g. G' (Figure 3), which is not galvanically connected to the inverter's electrical circuit (e.g. relative to the protective PE ground of the drive system). However, if the ground potential point G' is shorted with the negative DC voltage bus, then according to the Figure 3 the voltage $u_{GG'} = U_d/2$ and then the following applies 9 [11]:

$$u_{NG'} = u_{NG} + u_{MG'} = u_{NG} + \frac{U_d}{2} \tag{9}$$

Taking into account 9 and the voltages from the equation 6, a dependency considering the sum is obtained 10:

$$u_{NG'} = \frac{u_{AN} + u_{BN} + u_{CN}}{3} + \frac{1}{2}U_d \tag{10}$$

From 10 a conclusion can be drawn that the grounding of the negative DC voltage bus of the inverter does not eliminate the inverter's common-mode voltage $u_{NG'}$, but instead only increases its level relative to the ground. The grounding of the negative DC voltage bus of the inverter causes the ground disturbances current caused by this voltage, directly flow into the voltage inverter. It is a property caused a higher impedance in industrial frequency converters, because in this type of inverter's there is no galvanic connection between the negative DC voltage bus and ground [11].

4 Limiting Disturbances of High-Frequency Voltage Inverters in a Hybrid Grid

The DM and CM voltage of the PWM inverter are its incherent feature [12], [13]. The filtration of these undesired voltages eliminates a number of undesirable phenomena that occur when high-frequency capacitive leakage currents occur: phase-to-phase current from DM voltage and ground current from CM voltage. The use of significant inductance in the DM voltage filter, which simultaneously impedance separates the voltage inverter from the resonant current harmonic filter, simultaneously strengthens the level of current harmonic filtration.

The use of DM and CM voltage filtration is of fundamental importance for limiting the spread of high-frequency disturbances in the hybrid power network [12], [14]. High-frequency ground leakage capacitance currents usually flow through residual current circuit breakers without causing them to respond to an exceeded ground fault current, resulting in a risk of electric shock [15], [16]. Non-sinoidal phase voltages, e.g. in the form of a symmetrical rectangular waveform, will always generate a CM rectangular voltage. Rectangular phase voltage with phase shifts close to zero arise with the sinusoidal PWM modulation of the inverter and the modulation depth factor M is close to zero.

In the inverter model from 5 supplying the diode rectifier from Figure 10, a parallel resonance filter was used to limit harmonics in phase currents. Inductance (L3, L4, L5) of the DM voltage filter increases the impedance between the inverter and the resonant current filter, thereby increasing its efficiency [5], [13].

Suppression of the CM voltage harmonics of the PWM inverter is possible by forcing the capacitive leakage current CM as shown in Figure 5. When the AM1 ammeter current is zero, only the harmonics of the DM voltage are suppressed and the inverter phase-to-phase voltages are sinusoidal.

When the i_0 = AM1 current is different from zero then additionally the phase voltages are also sinusoidal and the CM voltage can be described by the formula 11:

$$u_{CM} = \frac{1}{3}\frac{1}{C_f}\int i_0 dt \tag{11}$$

where: i_0 = current at AM1, C_f - filter phase capacitance (Figure 5: C6, C7, C8 and if C5 $\gg C_f$ the $u_{CM} = u_{C5}$).

Formula 11 shows that with sufficiently large phase capacities C_f of the DM voltage filter it is possible to obtain phase voltages in the form of pure sinusoidal waveforms. The voltages obtained from the inverter (e.g. photovoltaic) supply the premises with a three-phase AC voltage line to which single- and three-phase receivers can be connected. Therefore it should be free of harmonics deforming phase voltages. In the present case, the AC line is loaded with the diode rectifier of the EV battery charger (Figure 10). The rectifier battery load models the load while charging the battery storage (e.g. DC/DC loaded converter). When galvanic connection of the inverter power outputs with the rectifier power inputs, a sufficiently large inductance value in the current path must be ensured. In the simulation tests, the inverter sinusoidal PWM modulation with a carrier frequency of 6kHz is used (modulation depth factor M = 1).

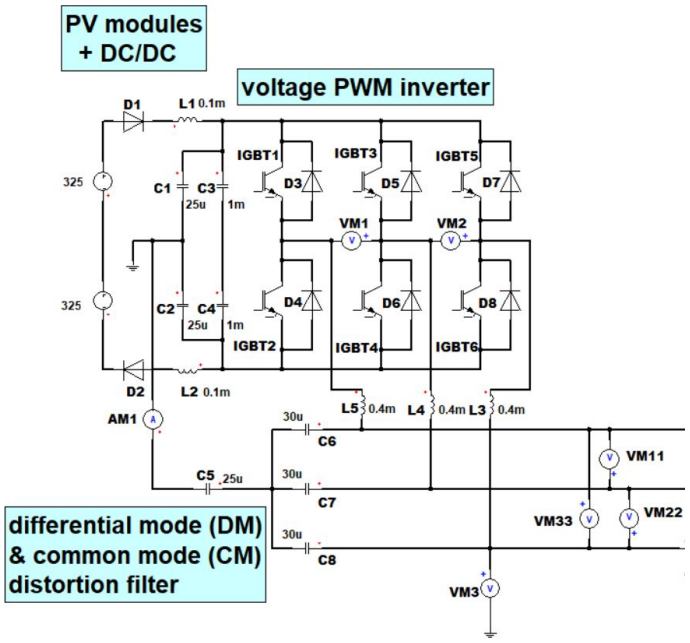


Figure 5: Two-level PWM inverter with DM and CM voltages filters with local AC line to supply of diode EV charger rectifier

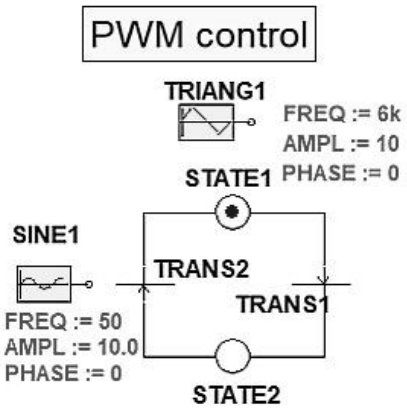


Figure 4: Event graph for PWM control of a three-phase inverter for one phase

The inverter PWM control in the form of an event graph is shown in the Figure 4. The carrier section is a symmetrical triangle with a frequency of 6kHz. Modulating waveforms are 50 Hz sine waves. In each phase they are phase shifted by an angle of 120°. For conducted tests of current harmonics produced in the rectifier and CM and DM inverter voltages, sinusoidal PWM modulation is used.

In the inactive phase of the inverter (i.e. when energy is not supplied by the inverter), the energy accumulated in the current path inductances maintains a sufficiently large rectifier supply current. The LC inductance of the DM disturbance filter accumulates energy (W_L=0.5LI^2), which supports the flow of phase currents to the rectifier when PWM modulation blocks the flow of currents through the inverter IGBTs.

The model of inverter with PWM sinusoidal modulation to feed diode rectifier in electric vehicles charging station is shown in the Figure 5. The reference [17] proposes other system structure with use the isolation transformer or boost converter [18] with inverter [1]. The inverter is powered by a solar power plant using PV modules (solar energy). The inherent DM and CM voltages of the PWM inverter are minimized with the DM filter (phase-to-phase harmonics) and CM filter (phase voltage harmonics after filtering the DM harmonics first) [6].

The PWM inverter model with DM and CM voltage filters used in simulation tests is shown in Figure 5. It enables simulation tests of the hybrid power system [5]. Ansys Simplorer software was used to build the model and perform simulation tests.

The phase-to-phase waveforms of the inverter obtained in simulation tests indicate that the parallel resonant harmonic filter in phase currents strengthens the inverter's DM voltage filtration efficiency. Comparing the phase-to-phase voltages of the inverter shown in Figures 6 and 7, a nearly 10% reduction in the THD_u harmonic content factor is visible when the current harmonic filter is on.

The voltmeters VM1 and VM11 used in the voltage inverter model from Figure 5 indicate the phase-to-phase voltages before and after the DM voltage filter, respectively. Figure 6 shows voltages when switched off and figure 7 when the resonant current harmonic filter is switched on.

Occurring surges in the phase-to-phase voltage lead to numerous undesirable phenomena, including resonance phenomena. Suppression of higher harmonics at phase-to-phase voltage is fundamental when limiting the high frequency disturbances spread in the local AC network. The beneficial effect of the applied harmonic phase current filter on increasing the harmonic suppression of the DM voltage in the phase-to-phase voltages of the inverter is shown in Figure 7.

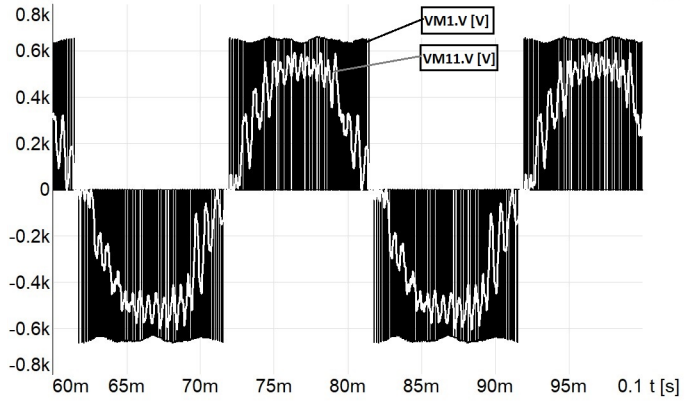


Figure 6: Phase-to-phase voltage of loaded PWM inverter (VM1) before DM and CM filters and after DM and CM filters (VM11 - THD_{1u}=17.55%) when resonance harmonic filter is **off**

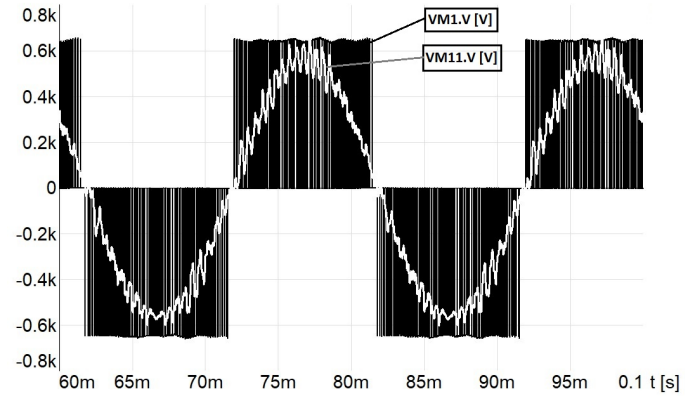


Figure 7: Phase-to-phase voltage of of loaded PWM inverter (VM1) before DM and CM filters and after DM and CM filters (VM11 - THD_{1u}=7.65%) when resonance harmonic filter is **on**

5 Model and Research of a Current Resonant Filter in a Three Phase Rectifier System Powered by a Voltage Inverter

Fast EV charging stations can be DC or AC powered. DC supply from renewable energy generators or electrochemical reservoirs (e.g. lithium-ion storage) relieves the local AC network. Power supply for the input rectifiers of EV battery chargers (DC/DC converters) requires harmonic current filtering, e.g. by using 5, 7 and 11 harmonic passive parallel filters. The harmonic filter is attached to a three-phase AC line depending on the rectifier's current rating. At low current diode rectifiers, the filter is disconnected from the AC line to limit the reactive capacitive power supplied to the power system (Figure 2).

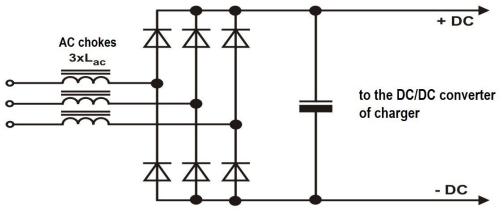


Figure 8: Chokes at the diode rectifier inputs to limit the steepness of the capacitor bank charging current pulses - AC chokes

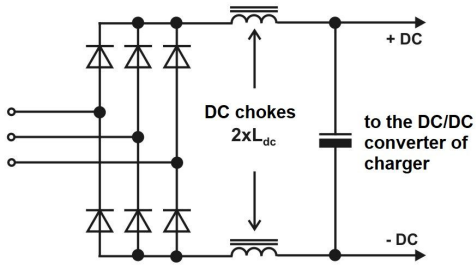


Figure 9: Chokes on the DC voltage side of the diode rectifier to limit the steepness of the current pulses of the capacitor bank charging - DC chokes

Three-phase diode rectifiers are equipped with AC chokes - Figure 8 or DC - Figure 9, or a combination thereof. These are usually chokes that cause a phase voltage drop from 3% to 4% at nominal current, which results in providing THD_i harmonic content ratio below 45% [1]. AC chokes protect the AC line against switching commutation of current in rectifying diodes, but they cause undesirable reduction of the rectified voltage value. DC chokes do not lower the straightened tension. If the source of AC voltage supplying the rectifier in the hybrid power supply network is a voltage inverter, then when determining the value of the AC or DC choke inductance of the rectifier, the inductance of the inverter's DM voltage filter should be taken into account. Usually the AC and DC chokes have similar inductance values, which decrease proportionally with increasing power of the rectifiers [5]. DC reactors are placed symmetrically in the branches $+U_{DC}$ and $-U_{DC}$ to ensure a symmetrical CM voltage of the rectifier relative to the ground potential, when it is fed from a power network with a TN network system.

The values of the capacity of the batteries and chokes used in the diode rectifiers for 50kW and 500kW EV converters are summarized in Table 1. The proportional relationship between the capacity of the EV rectifier batteries (figures 8 and 9) and its power is visible here, and the inductance of the chokes is inversely proportional.

Table 1: Basic parameters of EV rectifiers with power 50kW and 500kW

$P_r[kW]$	50	500
$C_{DC}[mF]$	10	100
$L_0(4\%) = L_{DC}[mH] = L_{AC}[mH]$	0.3	0.03

Using the L_0 AC or DC chokes in the rectifier with a C_{DC} capacitor battery EV charger with values from the table 1 reduces the harmonic content THD_i in the phase current of the AC line to a value of approx. 40%-45% [1]. Further reduction of the harmonic content coefficient, to the level of nearly 12% (Figure 12), was obtained by using a three-branch passive resonance filter (for 5th, 7th,

and 11th harmonic). Obtaining phase sinusoidal currents increases the efficiency of the EV charger rectifier supply system [6]. The AC and DC inductance of the chokes presented in table 1 take into account the inverter DM voltage filter phase inductance connected in series or the transformer inductance when the EV charger is powering from the power grid.

6 Model and Simulation Tests of The Rectifier's Harmonic Current Passive Filter

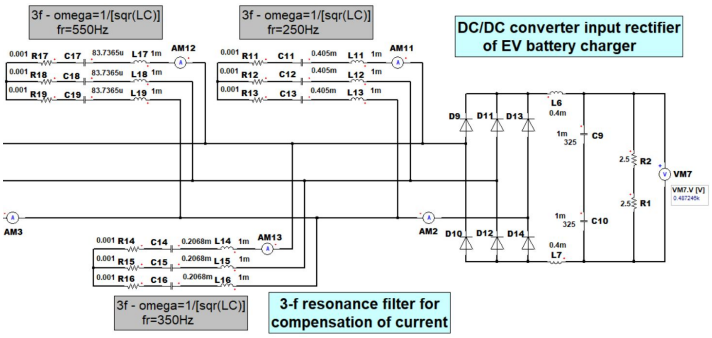


Figure 10: Model of 6-diode rectifier with parallel resonant filters for 5th, 7th and 11th harmonics

Diode rectifiers are commonly used to convert AC voltage to DC. In the DC/DC converter, the voltage is processed in such a way as to maintain a constant allowable current for charging the EV battery (Figure 1). In the rectifier model shown in Figure 10, the individual components mean: bank of capacitors (C9, C10), rectifier resistance load (R1 + R2), harmonic resonant filters for 5th, 7th and 11th harmonics.

Fast Fourier transform was used to set the coefficient of the THD_i phase harmonics. The results of simulation tests are shown in Figures 11 and 12.

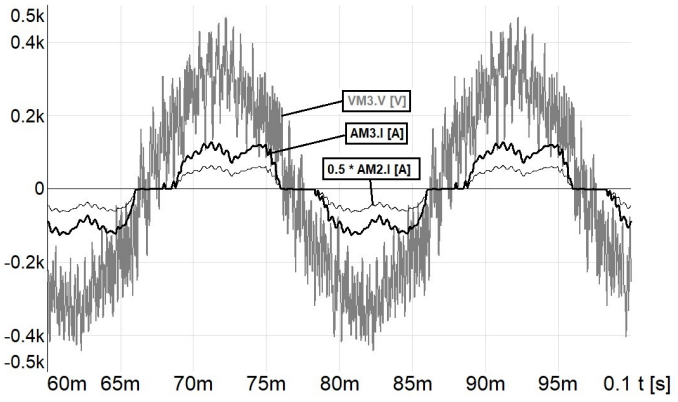


Figure 11: Phase voltage and currents waveforms on AC local network with the PWM inverter and diode rectifie: VM3 – phase voltage before rectifier, AM2 (THD_i=45,4%) current before the harmonic filter, AM3 - current after the harmonic filter when filter is off (Figure 10)

The phase current harmonics in Figures 11 and 12 are effectively filtered by means of a passive resonant filter. The high efficiency of the resonant filter is obtained by using the inductance of the inverter's DM voltage filter (Figure 5). In simulation studies, effective harmonic filtration has been demonstrated in the power supply system of a voltage inverter - diode rectifier, without the use of a separating transformer between these converters.

Figures 11 and 12 also show the VM3 phase voltage of PWM inverter loaded with active power of 50kW. The sinusoidal shape of the VM3 phase voltage indicates the effective operation of the CM voltage filter on PWM inverter outputs (Figure 5).

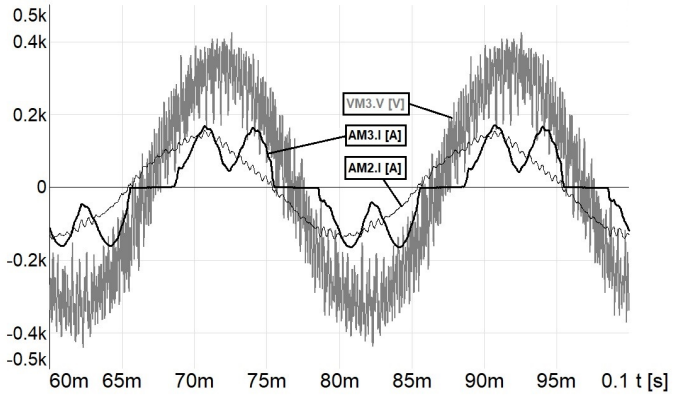


Figure 12: Phase voltage and currents waveforms on AC local network with the PWM inverter and diode rectifier: VM3 – phase voltage before rectifier, AM2 - current before the harmonic filter, AM3 (THD_i=12.21%) - current after the harmonic filter when filter is on (Figure 10)

7 Laboratory Experiments Results

Laboratory stand to testing cooperation of PWM volatge inverter (1) with rectifier (7) of drive inverter loaded with motor (11) is presented on the Figure 13.

The stand was built to examine the cooperation between a voltage inverter (for example, as a voltage source of a photovoltaic power plant) and a diode rectifier loaded with capacitive battery (for example, as an input stage of the DC/DC converter of an electrochemical energy storage charger). This solution does not use a separating transformer between the inverter and the rectifier, because it is an expensive solution, with large dimensions and weight. The authors are looking for an alternative (low cost) solution without using transformer separation between the inverter and the rectifier.

The main purpose of the research is to determine the conditions for cooperation of a PWM voltage converter generating three-phase alternating voltage with a DC/DC converter for charging EV batteries as shown in the Figure 1.

The laboratory stand, depicted on the Figure 13, for AC nominal line currents 3x10A/50Hz and phase-phase voltages 3x400V/50Hz contains:

- LC filter (3xL=4mH, 3xC=3μF) at the inverter output (differential-mode voltage filter) (4), which limits high frequency disturbances in phase-phase voltages of inverter (1),
- radio frequency interferences filter RFI (2) of the inverter (1),
- branch of CM voltage filter: 2xC=2x15μF, L=1mH (5),(6),

- diode rectifer of the voltage inverter (7),
- excess energy dissipation resistor (8),
- load system (11),(12) of the voltage inverter (7).

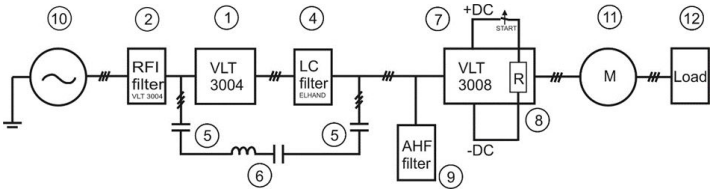


Figure 13: Block diagram of the laboratory stand for testing the cooperation of the renewable energy voltage inverter with the diode rectifier of the vehicle battery charging system

Figure 14 shows the measured line current and phase voltage on the power supply of the loaded diode rectifier of AC/DC/AC converter with 10A current. A significant deformation of the line current in the measured phase is visible as a result of charging the rectifer's (AC/DC converter) capacitor banks (THD_i approx. 45%).

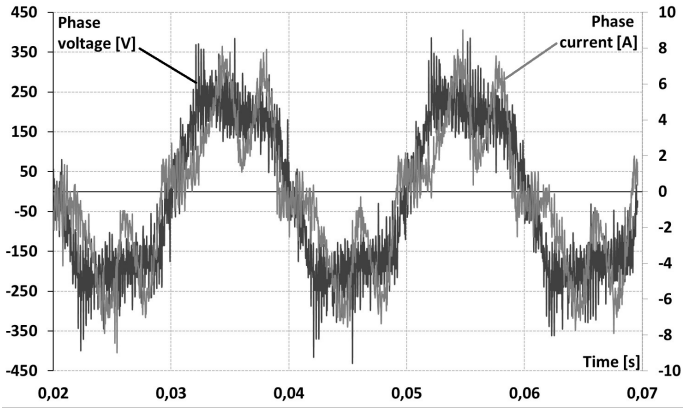


Figure 14: Phase voltage and phase current after the LC filter (4) at the test stand shown in the Figure 13 with the passive harmonic filter AHF [order: 5, 7, 11] switched **off** (9)

The phase voltage has a limited harmonic content of high orders, which confirms the correct operation of the CM filter (Figure 13). Previous simulation tests of the model of the local power supply system for vehicle battery charging stations are consistent with the experimental results (Figure 11) The capacitive reactive power consumption is not significant here,which results from a negligible phase shift of the first harmonic of the current and phase voltage.

The increased consumption of capacitive reactive power is observed after connecting the AHF passive harmonic filter (orders: 5, 7, 11), also under nominal load conditions, as shown in the oscillogram from the Figure 15 and the values are $\cos\phi_1$=0.89, ϕ_1=27°.

The oscillogram from the Figure 15 confirms the effective filtration of low order harmonics through the attached parallel resonant hamonic filter. The high efficiency of the LC passive filter was obtained by using nearly 3% current path inductance (i.e. 3% phase voltage drop on the inductance at nominal current), such as rectifier DC inductance (Figure 9) and DM disturbance filter inductance (Figure 5). Oscillogram from the Figure 15 also shows that the use of a resonant parallel harmonic filter has enabled the reduction of phase current harmonics to THD_i = 12.2% from above 45%. The experimental result confirmed the previously obtained simulation test result of the model shown in the Figure 12.

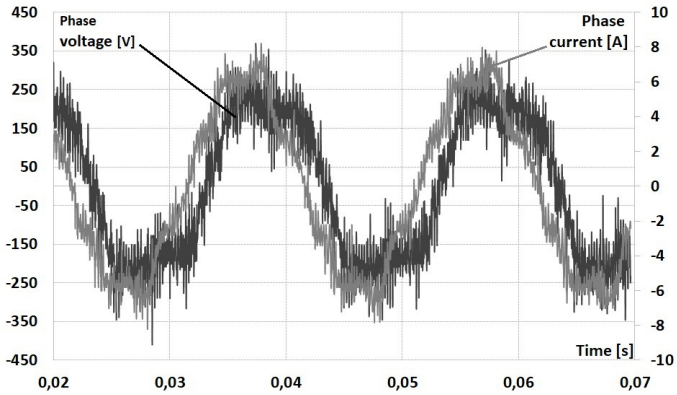


Figure 15: Phase voltage and phase current after the LC filter (4) at the test stand shown in the Figure 13 with the AHF passive harmonic filter [order: 5, 7, 11] switched **on** (9)

8 Conclusions

The presented system from transformerless supply of a 6-diode rectifier from a two-level PWM inverter is a low cost alternative to supplying the rectifier using a transformer. The tested power supply system can be used in fast EV charging stations. The system enables building a scalable EV charging station composed of modules, e.g. 50kW each.

In the tested system, the current efficiency of the inverter is comparable to the load power of the rectifier.

The applied CM and DM voltage filtering of the inverter made it possible to obtain practically sinusoidal phase and phase-to-phase voltages in the local three-phase AC network.

By using a passive resonant current harmonic filter, practically sinusoidal phase currents in the AC network were obtained at high loads of the diode rectifier, which increases the efficiency of the tested power supply system.

The applied DM differential voltage filters and CM common voltage provide a sinusoidal shape of phase and phase-to-phase voltage in the three-phase local AC network. The inductance of the 3xL DM filter also allows the accumulation of energy necessary to maintain the rectifier current continuity. In the inactive phase of the inverter (operation with passive control vectors: 000 or 111), the energy accumulated in the DM filter inductances is transferred to the rectifier.

To obtain sinusoidal phase and phase-to-phase voltages in the local hybrid three-phase voltage network, it is necessary to use DM and CM voltage filters of the inverter. The inductance of the DM voltage filter enhances the efficiency of the resonant harmonic filter of phase currents. If the CM filter is using, an additional energy storage to limit the voltage rise above the permissible value in the inverter's DC circuit is need to consider. In case the PWM voltage inverters feed the rectifier circuits, it is necessary to use passive or active harmonic filter. The use of a passive harmonic current filter

between the inverter and the rectifier has made it possible to reduce the THD_i of phase currents to a level similar to that of 12-pulse rectifiers with a transformer.

Simulation and experimental tests performed of the hybrid network model using the PWM inverter, filters and the rectifier showed the possibility of obtaining three-phase voltage source of PWM inverters with limited harmonic content.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgments In the part of the article related to the modelling of the electric circuit, the ANSYS Software national scientific software license has been used, which was funded by a computational grant obtained by Kazimierz Pulaski University of Technology and Humanities in Radom, Poland.

References

[1] Hansen S., Asiminoaei L., Blaabjerg F., Simple and Advanced Methods for Calculating Six-Pulse Diode Rectifier Line-Side Harmonics, 38th IAS Annual Meeting. Conference Record of the Industry Applications, vol. 3, 2003, pp. 2056 – 2062, DOI:10.1109/IAS.2003.1257849.

[2] Szymanski J., Common and differential mode voltage in frequency conventer drive systems. ElektroInfo, vol. 11, 2013. (article in Polish with an abstract in English).

[3] Li X., Xing X., Zhang Ch., Chen A., Qin Ch., Zhang G., Simultaneous Common-Mode Resonance Circulating Current and Leakage Current Suppression for Transformerless Three-Level T-Type PV Inverter System, IEEE Transactions on Industrial Electronics, Volume: 66 , Issue: 6 , June 2019, DOI: 10.1109/TIE.2018.2860555.

[4] Trivedi N., Gujar N., Sarkar S., Pundir S.P.S., Different fast charging methods and topologies for EV charging, 2018 IEEMA Engineer Infinite Conference, DOI: 10.1109/ETECHNXT.2018.8385313

[5] Szymanski J., Zurek-Mortka M., Harmonic Resonant Filters of Quick Battery Charging Station of Motor Vehicles, 23rd International Conference Electronics, IEEE, 2019, DOI: 10.1109/ELECTRONICS.2019.8765691

[6] Zurek-Mortka M., Szymanski J.: Reactive power compensation in a harmonic environment. elektro.info, vol. 10, 2018. (article in Polish with an abstract in English).

[7] Pedzisz K. Capacitor banks measuring analysis and operating condition assessment in industrial environment, Journal Electrical engineering and electronics, vol. 27, p.2, 2008.

[8] Kempski A., Smoleński R., Decomposition of EMI Noise Into Common and Directional Modes in PWM Inverter Drive System. Electrical Power Quality and Utilization, 12 (no. 1), 2006.

[9] Pairodamonchai P., Sangwongwanich S., Exact Common-Mode And Directional-Mode Equivalent Circuits of In-verters in Motor Drive Systems Taking Into Account Input Rectifiers. IEEE PEDS'11, Singapore, 5-8 December, 2011.

[10] Ambrožič, V., Zajec P., Električni Servo Pogoni, Slovensko združenje elektroenergetikov CIGRE-CIRED, Ljubljana 2016.

[11] Szymanski J., The Voltage of Common-Mode Disturbances in Frequency Converters in Railway Systems. Monography, UTH Radom, 2015.

[12] Tallam R. M., Skibinski G. L., Shudarek T. A., Lukaszewski R. A. Common-mode voltage reduction PWM algorithm for AC Drives, IEEE Transactions on Industry Applications, vol. 46, no. 5, pp. 1959–1969, 2010. DOI: 10.1109/TIA.2010.2057396.

[13] Takahashi S., Ogasawara S., Takemoto M., Orikawa K., Tamate M., Common-Mode Voltage Attenuation of an Active Common-Mode Filter in a Motor Drive System Fed by a PWM Inverter, IEEE Transactions on Industry Applications, pp.1-1, 2019. DOI: 10.1109/ICEMS.2017.8056376.

[14] Danfoss, Output filters design guide - MG90N402, 2010.

[15] Czapp S. Operation of residual current devices at an increased frequency of differential current. ElektroInfo, vol.10, 2008. (article in Polish with an abstract in English).

[16] Charoy A., Dunand Ph., Bearing currents induced by a power drive. Automotive Power Electronics, 26 - 27 Sept., Paris, pp.1-7, 2007.

[17] Park J.K., Kim Y.H., Jung Y.Ch., Won Ch.Y., A novel control strategy of an active clamped flyback inverter with synchronous rectifier for a photovoltaic AC module system, 8th International Conference on Power Electronics – ECCE Asia, 2011. DOI: 10.1109/ICPE.2011.5944557.

[18] Genikomaskis K.N., Gutierrez I.A., Thomas D., Ioakimidis C.S., Simulation and design of fast-charging infrastructure for a university-based e-Carshring System, IEEE Transactions of Intelligent Transportation Systems, vol.19, no.9, 2018. pp. 2923–2932. DOI: 10.1109/TITS.2017.2767779

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 302-306 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

The Impact of Manufactured Sand (M-Sand) as Partially and Fully Replacement of Fine Aggregate in Concrete

Altamashuddinkhan Nadimalla[1,*], Siti Aliyyah Masjuki[1], Siti Asmahani Saad[1], Ean Lee Woen[2], Siti Maisarah Ali[1], Naseer Ulla[3]

[1]*Civil Engineering Department, International Islamic University Malaysia, 53100, Malaysia*

[2]*Institute of Sustainable Energy, Universiti Tenaga Nasional, 43000, Malaysia*

[3]*Civil Engineering Department, Bearys Institute of Technology, 574153, India*

ARTICLE INFO

Article history:
Received: 23 November, 2019
Accepted: 19 January, 2019
Online: 27 January, 2020

Keywords:
flexural test
Impact test
compression test

ABSTRACT

This, research work was to study the potential of M-sand as compared to river sand in concrete, here M-sand is replaced by river sand 0%,45%,50%,55% and 100% in the concrete mix, Mix design is designed as per IS Standards. In this research a mix 1:2.32:2.82 (M20) was considered. The test specimen was casted for 7days, 28 days and 90days. The performance of M-sand was determined by several experiments such as slump test, impact strength test, flexural Strength, and compressive strength test. The results attained from each test states that as M-sand increases the slump value decreases. flexural strength, compressive strength and impact test of concrete at 7 days, 28 days and 90 days is greater at 100% and 50% replacement of M sand by river sand.

1. Introduction

This paper is an extension of work originally presented in conference IEEE 5th International Conference on Engineering Technologies & Applied Sciences, 22- 23 Nov 2018, Bangkok Thailand [1]. In the world, fine aggregate, probably natural sand, composes up to 30% of the volume of concrete, it is around 4 billion tons sand is required to meet annual necessity of concrete production [2]. Increasing in excavation of river sand from riverbeds leads to serious threat to environment [3]. Due to the limiting resources of river sand, alternatively M sand can be used. [1]–[10]. The Manufactured sand is produced by crushing the rocks [5] [10].

The impact of M-Sand on the concrete properties:

Higher split tensile strength, higher compressive strength, & higher flexural strength can be attained by the 50% substitution of of fine aggregate by M Sand [5], [6],[10]. It is due to the angular in shape of M sand gives good bonding between cement and aggregate. Gradual increase in strength can be obtained by 50% replacing river sand by M sand. [4]. Gained higher strength at 60% by substituting the river sand by M sand at various proportions such as 0%, 20%, 40%, 60% and 80% [14]. The concrete flexural strength increased up to 2% and 4.3% by 25% and 100% incorporation of river sand by M sand[15]. Compared to the air curing and standard moist curing, the membrane curing will give the good strength in both river sand and M sand. Addition of Super Absorbent Polymer leads to a significant increase of mechanical properties of the concrete [16].

The M sand properties are very much similar to river sand so nowadays, instead of river sand, M sand can be used in place of river sand[11] , by utilizing M sand in alternative of river sand the workability of concrete will decreases[12], the workability can be gained by adding water reducing admixtures [16].

It was observed that if river sand is completely replaced by M sand, M sand has greater resistance to loss in strength as compared to river sand, if specimen is immersed in chemicals. The Combined replacement of M sand and Marble powder upto 25% by river sand increased the concrete strength [17].

Compressive strength increased by approximately 10% for 28 days of curing, when river sand fully substituted by M sand [2]. It has been recorded that the compressive Strength enhance up to 50% incorporation of river sand by M-sand sand and introduction of 2% to 6% waste plastics [18].

The 100% incorporation of natural sand by M sand, it helps in increase in paste volume as compare to river sand, which is useful to produce self-compacting concrete. Increase in paste volume is

*Corresponding Author: Altamashuddinkhan Nadimalla, Email: altamashk1987@gmail.com

due presence of high fines in M sand, which increases the water demand also [19].

Manufactured sand is suitable as alternative for river sand at affordable cost. It will act as cohesive cement mortar. It helps environment to maintaining the economical balance [3] [19].

2. Objective

- To inspect the properties of M-sand
- To examine the impact of M-sand on fresh and hard properties of concrete.
- To investigate the comparison of the, flexural Strength, compressive strength and Impact sstrength achieved by the cubes and beams in river sand and M-sand.

3. Materials and Its Properties

3.1. Cement

In this research Ordinary Portland Cement (53 grade) was used. As per IS: 8112-1989 Cement Properties are determined as tabulated in Table 1.

Table 1: Cement Properties

Properties		Result	IS code Requirement
Specific gravity		3.05	
Standard consistency		31%	
Setting time	Initial(Minutes)	33 min	30 minimum
	Final(Minutes)	380 min	600 maximum
Compressive strength (MPa)	7 days	32	33
	28 days	46	43

3.2. Fine Aggregates

The river sand was collected from the Mangalore local area which is excavated from riverbeds. The sieve analysis of river sand is tabulated in Table 2 and shown in Figure 3.

Table 2: Sieve analysis of river sand

SL No	Sieve Designation	Percentage Passing	Grading Limits for Zone II Sand (IS383)
1	4.75 mm	99	90-100
2	2.36 mm	87.8	75-100
3	1.18 mm	54.6	55-90
4	600 micron	37.6	35.59
5	3000 micron	12.6	8-30
6	150 micron	6.4	0-10
7	75 micron	2	0-10

From the sieve analysis test the river sand and M-sand fineness value are 3 and 3.45 respectively, both falls under zone-2, it shows the M-sand properties is similar to river sand, M-sand is slightly coarser as compared to river sand. The sieve analysis of M sand is shown in Table 3 and Figure 4

Table 3: Sieve analysis of M- Sand

SL No	Sieve Designation	Percentage Passing	Grading Limits for Zone II Sand (IS383)
1	4.75 mm	100	90-100
2	2.36 mm	91	75-100
3	1.18 mm	61.2	55-90
4	600 micron	48.3	35.59
5	300 micron	27.4	8-30
6	150 micron	13.8	0-10
7	75 micron	3.8	0-10

3.3. Coarse Aggregate

The 20mm down size crushed stone is used in this research. Table 4 and Figure 5 shows the Coarse aggregate sieve analysis.

Table 4: Sieve analysis of Coarse Aggregate

SL No	Sieve Designation	Percentage Passing	Gradation requirement as per IS 383-1970 for grade II	Remark
1	40	100	100	As per IS 383 the sample confirms the graded aggregate
2	20	98	95-100	
3	10	46	25-50	
4	4.75	0.8	1-10	

3.4. Mix Design

In this research work a mix design 1:2.32:2.82 (M20) is considered.

Table 5: Mix Design

Mix Design	Cement (kg/m^3)	Aggregate (kg/m^3)		w/c Ratio	Water (l)
		Fine	Coarse		
1:2.32:2.82	358	829.67	1009.96	0.55	197

4. Fresh Properties of Concrete

4.1. Slump test

The workability of concrete was estimated by using slump test. Figure 6 and Table 6 shows that slightly decrease in the slump value as M sand increases.

Figure 1: Slump test

Table 6: Slump test

Percentage of M Sand	Slump Value (mm)
0%	102
45%	100
50%	99
55%	100
100%	99

5. Hard properties of Concrete

5.1. Compressive Strength test

The tests were carried out on 150mm* 150mm*150mm size of cubes. The outcomes is displayed in Figure 8.

Figure 8 shows the compressive strength for results for 7days, 28 days and 90 days curing, it is detected that the compressive strength increased for 100% and 55% of replacement of M Sand by river sand, 100% and 55% incoporation of M sand can be advised to use as fine aggregate.

Figure 2: Compressive strength test

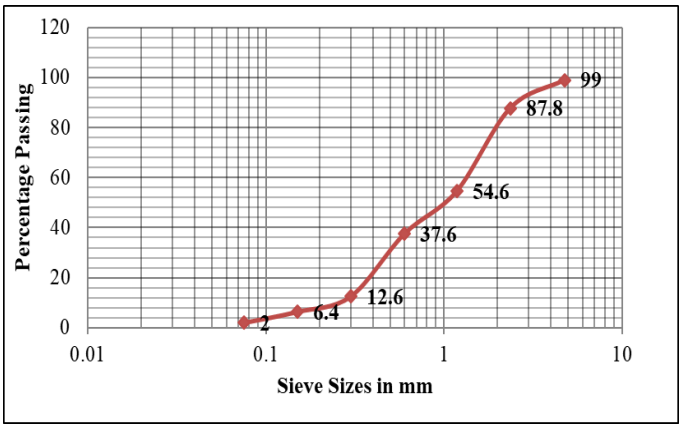


Figure 3: Sieve analysis of River Sand

5.2. Flexural strength test

The tests were carried out on beams of width 150 mm length 150 mm and 70 mm thickness. Figure 9 shows the flexural strength for results 7days, 28 days and 90 days curing, it is observed that the flexural strength increased for 100% and 55% of incorporation of M Sand by river sand, 100% and 55% incorporation of M sand can be advised to use as fine aggregate.

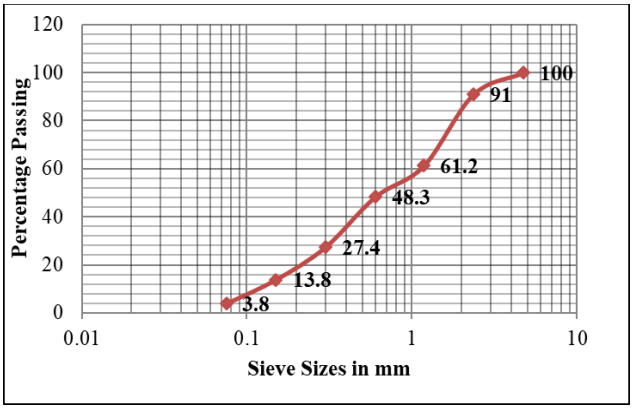
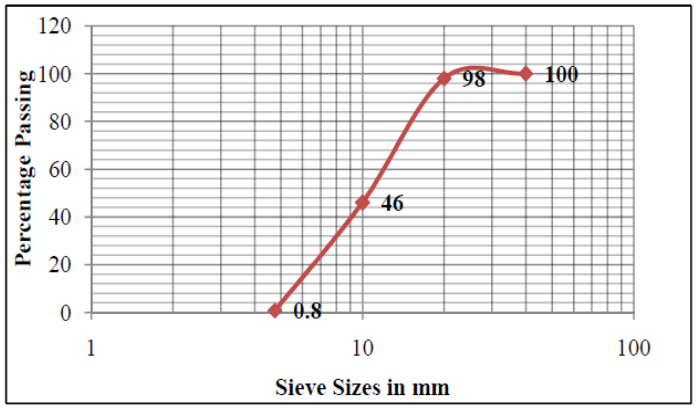


Figure 4: Sieve analysis of M Sand



5.3. Impact Test Results

Figure 10 shows the variation in impact test results of the concrete specimen prepared by replacing natural sand by M-sand at the percentages of 0%, 45%, 50%, 55% and 100%. Results shows that energy consumption in 55% and 100% replacement

shows increase in energy consumption that is about 47.99% and 40% for initial crack, 48.38% and 38.7% for ultimate failure when compare to 0% replacement. It can be concluded that 100% and 55% substitution of river sand by M sand has a threshold values for an optimal performance in energy adsorption and crack resistance.

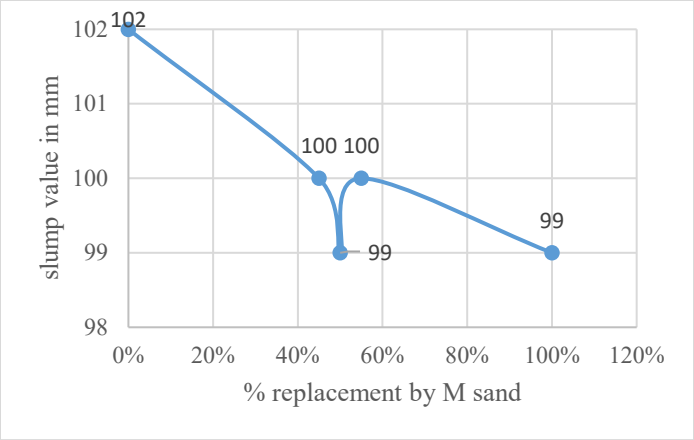


Figure 6: Slump test graph

Figure 6: Flexural strength test

Figure 7: Impact test on concrete

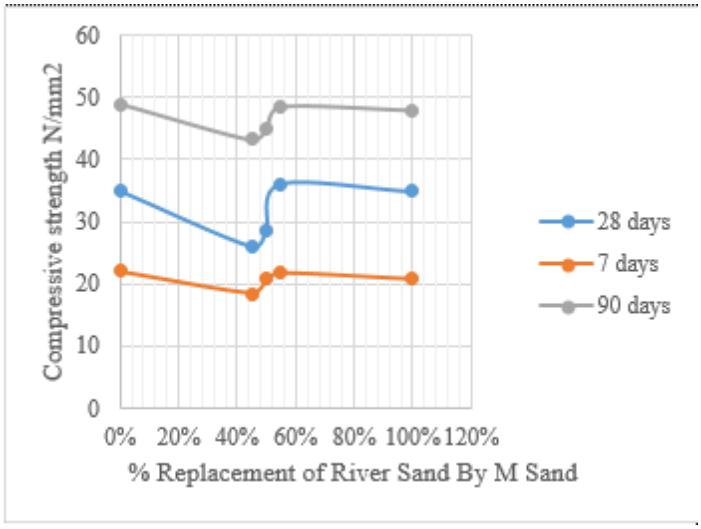
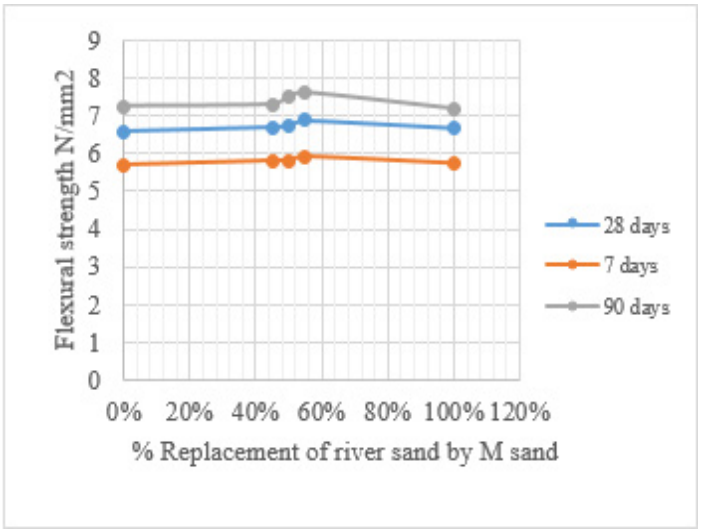


Figure 8: Compressive strength test graph



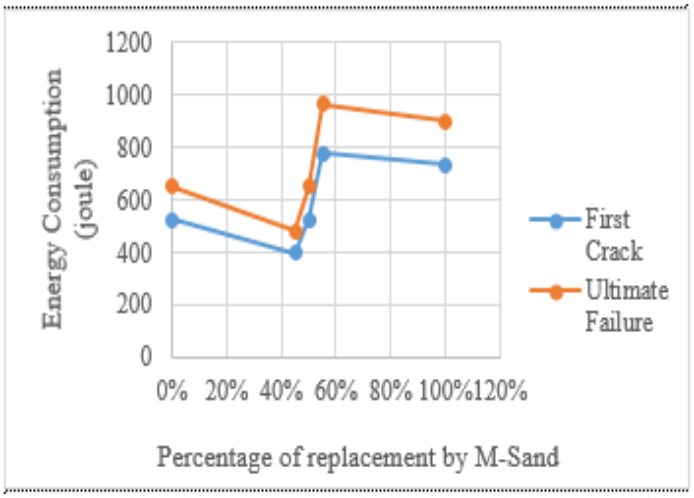


Figure 10: Variation in Energy Consumption for M20 grade Concrete Specimen

Conclusion

M-sand and River sand fineness value are 3 and 3.45 respectively, both falls under zone-2, it shows the M-sand

properties is similar to river sand, M-sand is slightly coarser as compared to river sand.

By 100% and 55% incorporation of M sand by river sand concrete can achieve higher flexural strength and compressive strength of concrete at 7 days, 28 days and 90 days.

Concrete can achieve optimal performance in energy adsorption and crack resistance by 100% and 55% substitution river sand by M-sand.

An incorporation of 100% and 55% of M sand can be advised to use as fine aggregate to enhance the strength of Concrete.

Acknowledgement

The authors would like to acknowledge financial support from iRMC Bold2025, Universiti Tenaga Nasional, Malaysia.Grant Code: RJO 1043 6494.

References

[1] Altamashuddunkhan, N., & Siti Alliyah, M, "The Effect of Replacement of Natural Sand by Manufactured Sand on the Properties of the Concrete. 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)", 22–23, 2018. DOI: 10.1109/ICETAS.2018.8629206

[2] A. Zimar, G. Samarawickrama, W. Karunarathna, and S. Jayakody, "Effect of Manufactured Sand as a Replacement for Fine Aggregates in Concrete," 8th Int. Conf. Struct. Constr. Manag., no. October 2017.

[3] P. A. Jadhav and D. K. Kulkarni, "Effect of replacement of natural sand by manufactured sand on the properties of cement mortar," vol. 3, no. 3, pp. 621–628, 2013.

[4] G. Nithyambigai and F. Ash, "Partial Replacement of Manufactured Sand and Fly Ash in Concrete," vol. 5, no. 6, 2015.

[5] K. U. Mani, N. Sathya, and R. Sakthivel, "International Journal of Modern Trends in Engineering and Research Effect of replacement of River sand by M-sand in high strength concrete," 2014.

[6] R. N. U. Shanmugapriya T., "Optimization of Partial Replacement of M-Sand By Natural Sand in High Performance Concrete With Silica Fume," Int. J. Eng. Sci. Emerg. Technol., vol. 2, no. July, pp. 73–80, 2012.

[7] N. R. Monika and S. Sarankokila, "Partial Replacement of M-Sand on Concrete by Self Curing and Standard Curing Method," pp. 12162–12169, 2015. DOI:10.15680/IJIRSET.2015.0412087

[8] T.Subramani & K.S.Ramesh, " Experimental Study on Partial Replacement of Cement With Fly Ash And Complete Replacement of Sand With M sand," Int. J. Appl. or Innov. Eng. Manag., vol. 4, no. 5, pp. 313–322, 2015.

[9] A. Jayaraman, "Experimental Study on Partial Replacement of Natural Sand with M- Sand and Cement with Lime Stone Powder," vol. 6, no. 2, pp. 948–954, 2014.

[10] T. F. Kala, P. Asha, and S. Elavenil, "Properties of Concrete using Manufactured Sand as Fine Aggregate," Int. J. ChemTech Res., vol. 11, no. 03, pp. 94–100, 2018. DOI: http://dx.doi.org/10.20902/IJCTR.2018.110337

[11] M. Meghashree, "Comparison of Physical Properties between Natural Sand and Manufactured Sand," vol. 3, no. 07, pp. 92–96, 2016.

[12] S. Suresh and J. Revathi, "Effect of M-Sand on Setting Time of High Performance Concrete," Asian J. Res. Soc. Sci. Humanit., vol. 6, no. 10, p. 1648, 2016. DOI:10.5958/2249-7315.2016.01118.7

[13] P. Jaishankar and K. Saravana Raja Mohan, "Experimental investigation on nano particles in high performance concrete," Int. J. ChemTech Res., vol. 8, no. 4, pp. 1666–1670, 2015.

[14] Y. Boopathi and J.Doraikkannan, "Study on M-Sand as a Partial Replacement of Fine Aggregate In Concrete," Int. J. Adv. Res. Trends Eng. Technol. vol. 3, no. 2, pp. 746–749, 2016.

[15] B. K. Meisuh, C. K. Kankam, and T. K. Buabin, "Effect of quarry rock dust on the flexural strength of concrete," Case Stud. Constr. Mater., vol. 8, no. July 2017, pp. 16–22, 2018. DOI: https://doi.org/10.1016/j.cscm.2017.12.002

[16] J. R. Sagura R, "Experimental Study on Mechanical Properties of M-Sand Concrete by Different Curing Methods.," IOSR J. Mech. Civ. Eng., vol. e-ISSN: 22, pp. 19–25.

[17] M. Idrees, " Utilization of Waste Quarry Dust And Marble Powder in Concrete," no. July, 2019.

[18] B. Bahoria, D. K. Parbat, P. B. Nagarnaik, and U. P. Waghe, "Effect of replacement of natural sand by quarry dust and waste plastic on compressive & split tensile strength of M20 concrete," Procedia Eng., no. March, p. 9, 2013.

[19] P. Nanthagopalan and M. Santhanam, "Fresh and hardened properties of self-compacting concrete produced with manufactured sand," Cem. Concr. Compos., vol. 33, no. 3, pp. 353–358, 2011.
DOI: https://doi.org/10.1016/j.cemconcomp.2010.11.005

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 307-316 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Effects of Split Ring Resonator (SRR) Metamaterial on the Radiation Pattern and Variation of the Heating Focus Point of the Microstrip Patch Antenna

Mahsa Jahangiri*, Saman Rajebi

Electrical Engineering Department, Seraj Higher Education Institute, Tabriz, Iran

ARTICLE INFO

Article history:
Received: 03 September, 2019
Accepted: 13 January, 2020
Online: 07 February, 2020

Keywords:
Microstrip patch antenna
Metamaterial
Split ring resonator
Circular hole
Hyperthermia
Specific absorption rate (SAR)

ABSTRACT

Metamaterial antennas are some sort of antennas which utilize metamaterials to improve the efficiency of small (small electrical) antenna systems and also enhance the antenna's radiation pattern. The unusual properties of metamaterials such as coefficients of permittivity and permeability, as well as the negative refractive index, have led to increase the use of these artificial materials in variant fields including microwave engineering and medical applications. This paper explores a metamaterial coating and impact of it on a microstrip patch antenna from two different perspectives on satellite and biomedical applications. For this purpose, a microstrip antenna is modeled using CST software and then a particular metamaterial cover is configured upon the patch antenna to enhance its efficiency such as radiation pattern and directivity. The method is arranged in several stages and the output of the antenna is then analyzed in each stage after applying some adjustments to the design process. On the other hand, because the type and position of the applicator on the tissue is a consequential variable, in order to improve the effectiveness of hyperthermia treatment in medical applications, in the next step the metamaterial antenna is applied to hyperthermia treatment on the body tissue. Eventually the results of metamaterial antenna are compared with the usual antenna.

1. Introduction

Current paper is an extended version of work originally presented in [1].

Veselago first proposed the concept of the negative refractive index and the presence of substances with this feature in 1968 [2]. By using a wired structure John Pandari succeeded in achieving negative permittivity in 1996, and by using split resonator rings, he then prospered independently in generating an ambit with negative permeability in 1999 [3]. The idea was fulfilled experimentally and actually in 2000 by David Smith and his fellow members, applying an intermittent volumetric series of alternative metal wires and split ring resonator combinations [4,5]. The name given to these materials is "Metamaterial".

Since the metamaterials have the negative ε and μ, n has a negative root in certain frequencies which is formulated by Maxwell equation that is $n = \pm\sqrt{\varepsilon\mu}$ [6]. The peculiar and appealing features of these materials have contributed to their usage in diverse fields, such as microwave engineering, waveguide operation, phase compensation, dispersion correction, smart new antennas, lenses and numerous other instances. A type of metamaterial structure and its impacts on performance of a microstrip antenna will be studied in this article.

If an electrical field is applied in parallel to a structure composed of a series of thin metal rods, a negative permittivity property will be revealed in the structure. Likewise, if a magnetic field is perpendicularly exerted to a mechanism with a set of split ring resonators (SRR), it will exhibit a negative permeability [7].

The Microstrip Antenna Concept was first presented in 1950 [8]. The reasons for its success were proper characteristics including lower volume, light weight, low expense and reasonable efficiency [9,10]. A plan is proposed in this paper that using it we will explore the capacities as well as restrictions of metamaterial mechanism in antenna design.

The standard microstrip antenna is formed of three segments of ground, a radiation surface and, a dielectric substrate. This

*Corresponding Author: Mahsa Jahangiri, Seraj Higher Education Institute, Tabriz, Iran, Email: jahangiri.mahsa66@gmail.com

https://dx.doi.org/10.25046/aj050139

dielectric substratum lies between the patch (radiation surface) and the ground layer.

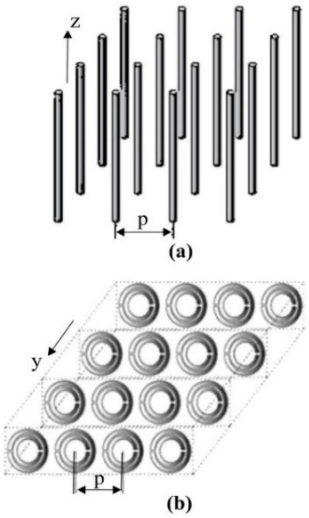


Figure 1: An illustration of intermittent patterns. (a) The metal thin rods, (b) The split ring resonators (SRRs)

Equations (1), (2), (3), (4) and (5) measure the sizes of a standard patch, where W and L are width and length of the radiation patch, respectively [9,11].

$$\varepsilon_{reff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r + 1}{2}\left[1 + 12\frac{h}{W}\right]^{-\frac{1}{2}} \quad (1)$$

$$W = \frac{c}{2f_0\sqrt{\frac{(\varepsilon_r + 1)}{2}}} \quad (2)$$

$$L_{eff} = \frac{c}{2f_0\sqrt{\varepsilon_{reff}}} \quad (3)$$

$$\Delta L = 0.412h\frac{(\varepsilon_{reff} + 0.3)(\frac{W}{h} + 0.264)}{(\varepsilon_{reff} - 0.258)(\frac{W}{h} + 0.8)} \quad (4)$$

$$L = L_{eff} - 2\Delta L \quad (5)$$

The theory of the effective quantity has been used here. We need data of the scattering parameters (S_{11} and S_{21}) for a wave incident normally on a finite slab of a homogeneous material, to recover its electromagnetic characteristics including analytical quantities of permeability (μ) and permittivity (ε), that express a straightforward description. S_{11} and S_{21} are measured via (6) and (7) [12]:

$$S_{11} = \frac{i}{2}\left(\frac{1}{z} - z\right) sin(nkd)\, S_{21} \quad (6)$$

$$S_{21} = \frac{1}{cos(nkd) - \frac{i}{2}(z + \frac{1}{z})\, sin(nkd)} \quad (7)$$

Also the permeability and permittivity are computed by the relations μ = nz and ε = n / z [12, 13].

As it is evident, scattering parameters are correlated with ε and μ, and also with n and z. Hence, by inverting the S-parameters, the refractive index (n) and impedance (z) would be derived. For the calculation of n and z, equations (8) [13] and (9) are used [14]:

$$z = \sqrt{\frac{(1 + s_{11})^2 + {s_{21}}^2}{(1 - s_{11})^2 - {s_{21}}^2}} \quad (8)$$

$$n = \frac{1}{kd}cos^{-1}\left[\frac{1}{2s_{21}}(1 - {s_{11}}^2 + {s_{21}}^2)\right] \quad (9)$$

In which $k=\frac{w}{c}$ indicates the wave number of free space and d is the slab thickness [12].

1.1 Overview of hyperthermia

Cancer rates are increasing alarmingly all around the world, and this matter is a major cause of concern for the medical community [15]. Cancer tumors are able to affect different tissues and spread to different body organs. In recent years, in order to deal with the disease, researchers have developed a variety of treatments, including surgery, chemotherapy, radiotherapy, immunotherapy, stem cell transplant, and targeted therapies. Beside these methods, another well-known method called hyperthermia (or heat treatment) has recently been evaluated in clinical trials. Hyperthermia is a treatment method in which tumor tissues are exposed to high temperatures to be eliminated, without affecting healthy tissues [16]. Hyperthermia can be performed in both invasive and non-invasive methods. The non-invasive hyperthermia treatment is an efficacious safe treatment with less pain and injury to counteract different types of cancer [17]. One of the biggest challenges in the non-invasive treatment of the hyperthermia, is the concentration of electromagnetic energy on the cancerous tissue in a way that not to damage to healthy tissues. In this treatment, it is required that the temperature of the treated area be raised from 35 ° C to about 42 to 45 ° C [16]. During the period of hyperthermia treatment, tumor cells are exposed to a temperature of about 42 ° C for an hour or even longer [15]. Hence, many researchers have recently focused on non-invasive hyperthermia antennas and applicators. The applicator has been considered as the principal part of the hyperthermia which uses electromagnetic energy to be transmitted and distributed in the form of heat. It is then absorbed by cancer tissues to a certain level of heat and time [17].

1.1.1 SAR in hyperthermia

The quantity that is investigated in hyperthermia topic, is a parameter called specific absorption rate (SAR). The SAR quantity represents the quantity of heat energy absorbed by living tissue per unit time, when the rise of body temperature due to an external heat source exceeds one degree celsius. Specific absorption rate is defined as the energy absorbed per unit mass of tissue in watts per kilogram by (10):

$$SAR = \frac{d}{dt}\frac{dw}{dm} = \frac{d}{dt}\frac{dw}{\rho dv} \quad (10)$$

The SAR measurement can also be expressed in terms of the tissue temperature rise by (11):

$$SAR = C_h \frac{dT}{dt} \quad (11)$$

The SAR can also be calculated as (12):

$$SAR = \frac{\sigma E^2}{\rho} \quad (12)$$

which in the (10), (11) and (12), E is intensity of the electric field in the tissue with unit of volt/meter, σ is conductivity of tissue liquid with unit of Siemens per meter, ρ is volumetric density of tissue with unit of kg/m3, C_h is specific heat capacity of tissue with unit of J/kg and Kelvin, dT/dt is primary time derivative of tissue temperature in Kelvin unit. Therefore, SAR is calculated by measuring the amplitude of the electric field intensity [18].

2. Methodology

2.1. Original thought behind current study

This article's original concept was taken from [19]. A microstrip patch antenna was proposed in the aforesaid paper, the radiation pattern was then ameliorated with the help of a metamaterial coating composed of split ring resonators (SRR) and wires that generate negative permeability and permittivity. Table 1 represents the findings of the above-mentioned paper.

It is worth mentioning that the SRR (Split-Ring Resonator) unit cell structure was derived from [20]. The distinction between them is that the unit cell was regarded as a singular square in the mentioned article while it is constructed from a couple of squares adjacent to one another in the present study.

Table 1: Compared results in paper [19] between the usual patch antenna and the metamaterial antenna

Parameters	Usual patch antenna	Metamaterial antenna
Return losses (dB)	-30	-23
Gain(dB)	6.80	7.32
HPBW(3dB)	80°	76°

2.2. Stages of configuration of the metamaterial antenna

To obtain the desired frequency in the 8-12 GHz range, the patch measurements (L × W) are calculated as (11.86 mm × 9.06 mm) and for the ground sizes (Lg × Wg) are measured as (21.34 mm × 18.54 mm). The frequency of resonance for the proposed antenna is 9.6 GHz [1].

The coaxial cable has been considered as the feeding of the aimed Microstrip patch antenna. In order to properly match with the input impedance, coaxial feeding can be positioned within the patch at any point [10]. To determine the location of the coaxial cable precisely, the input impedance is adjusted to: $cos^2(\frac{\pi y_0}{L})$ by 50 ohms.

Figure 2 displays a general form of an ordinary rectangular patch antenna which is stimulated by coaxial line.

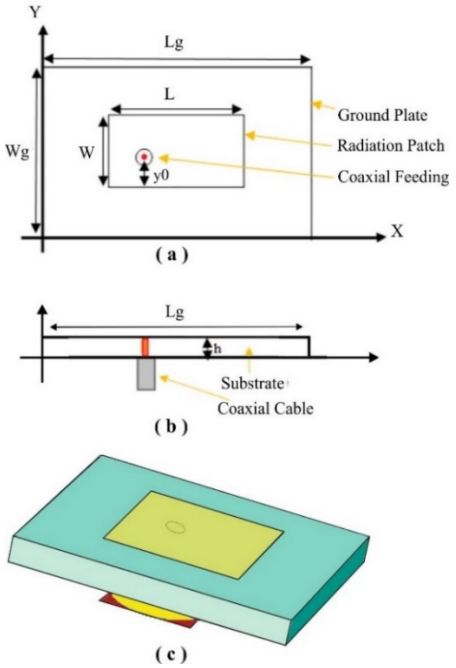


Figure 2: The microstrip patch antenna designed for 9.6 GHz. (a) Top-view geometry, (b) Side-view geometry, (c) Software-designed antenna

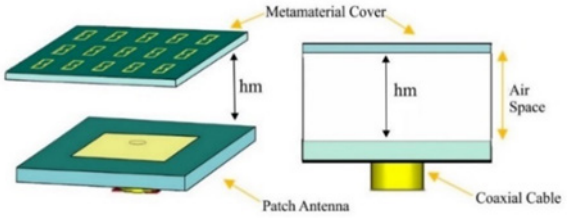


Figure 3: The metamaterial-coated microstrip patch antenna

The dielectric segment of the antenna has been constructed of Epsilon 2.2 substratum of Teflon and 1.58 mm thickness. Copper has been used to make the ground plane and radiation patch. In this paper, a cover of metamaterial has been utilized over a patch antenna as shown in Figure 3. This design forms of a 0.5 mm Teflon sheet of epsilon 2.2, which has 5*3 periodic slotted rings printed on the surface of this Teflon coating (single layer) measuring 0.017 mm thick and built of copper. This system consists of simple elements which are extremely expandable. The metamaterial layer is positioned from the patch of antenna at a distance of hm=10.53 mm, which is an air gap. This air distance was assumed to be zero at first, and the metamaterial was located directly on the antenna. Then in order to compare the changes, an air distance between the antenna and the metamaterial was regarded. The comparisons and findings have been provided in the section 3. In another phase, the SRR's intermittent configuration was arranged in two layers which were mounted on both sides of the Teflon coating surfaces in equal numbers. The findings have been analyzed in section 3 and compared to the singular layer SRR type. During other stage, we modified the unit cells and attained new results. At the end, the function of metamaterial antenna in hyperthermia treatment, has been investigated.

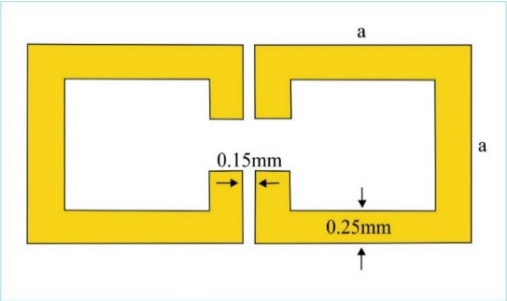


Figure 4: The designed SRR unit cell

Figure 3 displays the common microstrip antenna with the placement of metamaterial cover and Figure 4 illustrates the unit cell of SRR. As shown in Figure 4, there are two square-like components of the unit cell with a distance between them that its size is 0.15 mm. Line width of each component is 0.25 mm and the length is a=1.46 mm [1].

During the first phase, the outcomes of the antenna without metamaterial coating have been explored, taking into account the measurements and configuration of the coaxial feeding microstrip patch antenna.

In general, the real length of the patch in this group of microstrip antennas is the half of the wavelength, it has a weak directivity of almost 5 to 6 dB and a half-power bandwidth scope of around 70 to 90 degrees [21]. In this study, the purpose of applying metamaterial cover is to boost gain and beam direction. The aimed spectrum has been planned for high-frequency programs ranging from 8 to 12 GHz.

2.3. Design of metamaterial antenna on the body tissue in the hyperthermia application

At this step, three layer of body tissue including skin, fat and muscle in both simple (no metamaterial) mode, and then with metamaterial cover, is posited on the microstrip patch antenna. Figure 5 shows a metamaterial antenna with tissue.

In Figure 5, a skin layer of 1 mm thickness, a fat layer of 10 mm thickness, and a muscle layer of 40 mm thickness at a distance of Ag=25 mm from the microstrip patch antenna - which is actually an air gap - have been placed. Our aim is to investigate the effect of this air distance on the amount of output SAR. For this purpose, we have placed metamaterial coating at different distances from the tissue (Ag-hm) and then measured the amount of output SAR. As it is shown in Figure 5, by changing the Ag-hm distance, the gap between the metamaterial cover and the patch antenna (hm) will also change. Because, as mentioned, the overall air gap between the antenna and the tissue must remain constant (25mm).

3. Results and discussion

Figure 6 displays the s-parameter features of the metamaterial-free antenna.

It can be understood that the resonance frequency exists at about 9.63 GHz and that the S_{11} at this frequency is roughly -41 dB which has a reasonable value.

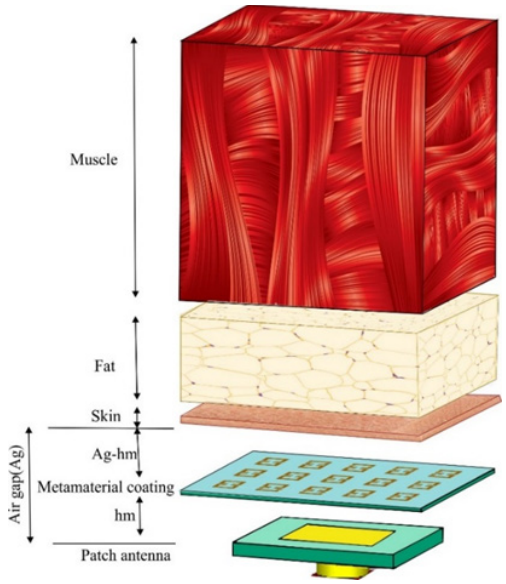


Figure 5: Geometry of the presented metamaterial antenna with body tissue.

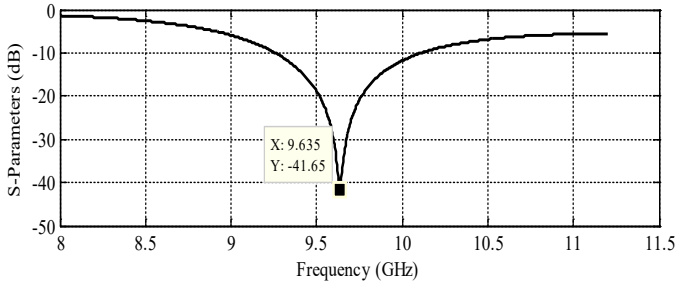


Figure 6: The antenna return losses without the use of metamaterial cover

The value of S_{11} in Figure 6, refers to the most power being transmitted to the antenna, and that is one of the fundamental definitions in the subjects of antennas. Figure 7 displays the pattern of radiation and the far field in metamaterial-free antenna. As shown in Figure 7, the gain is 7.66 dB at a frequency of 9.6 GHz. Half power beamwidth (HPBW) is the angle between two main beam points, 3-dB below the maximum power. That is 78.9 degrees here, and the main lobe direction is zero. Reducing the HPBW value is our aim for high-frequency applications. Because it indicates a rational directivity for our antenna [1].

3.1. Effect of airspace between antenna and metamaterial

In this section, there is not any gap between the metamaterial coverage and the patch of antenna, and the metamaterial coating including 15 SRR cells printed on a Teflon sheet has been directly mounted on it. In this case, Figure 8 displays the return losses of the antenna, and Figure 9 illustrates the return losses when 10.53 mm of air distance has been applied to the system.

In this situation, frequency has shifted roughly 0.7 GHz than the antenna without metamaterial. The metamaterial coverage is thus positioned at an optimal air gap of 10.53 mm as better results have been derived from the presence of airspace. According to Figure 9, the return loss of the patch antenna is about -50 dB due to the metamaterial coating structure which grants a remarkable performance. Also, the resonance frequency did not change substantially than the base antenna.

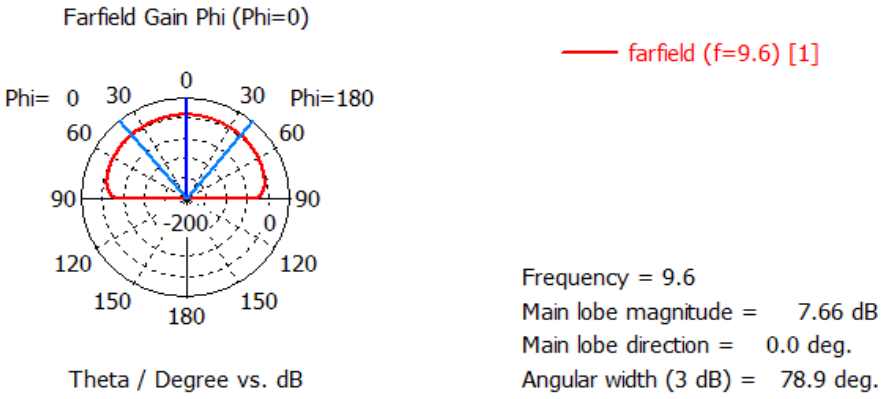


Figure 7: The radiation pattern of the metamaterial-free antenna

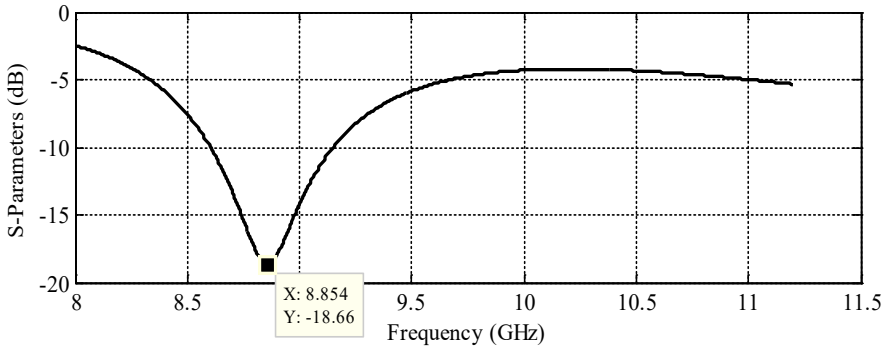


Figure 8: The return losses of antenna without airgap

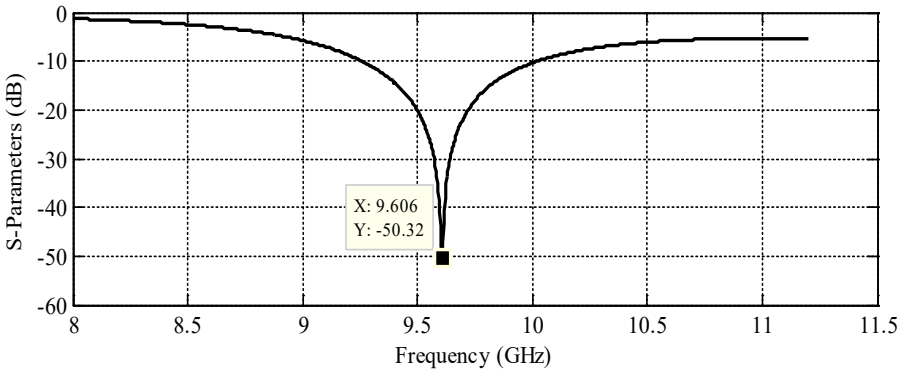


Figure 9: The return losses with airspace of 10.53 mm between the metamaterial and the antenna

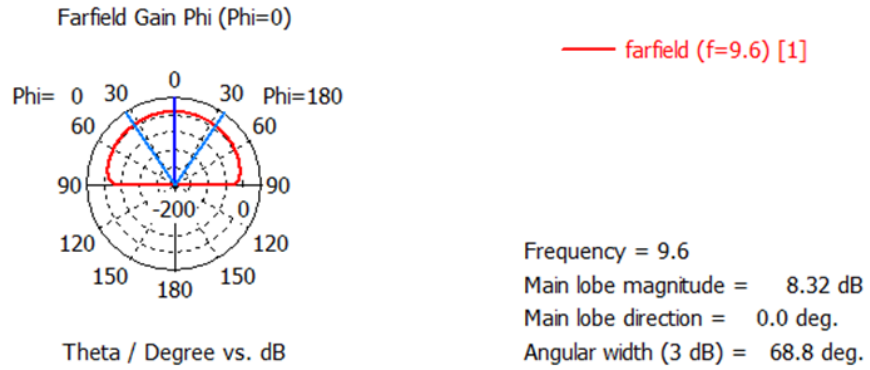


Figure 10: The radiation pattern of the metamaterial-coated antenna

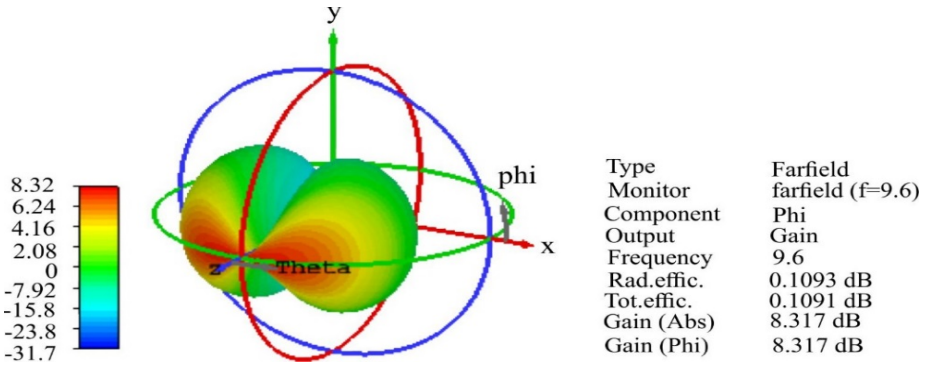


Figure 11: Three-dimensional view of the radiation pattern

Table 2: Comparison of outcomes between antenna without and with metamaterial (MM)

Parameters	MM-free Antenna	MM-coated Antenna
Return losses (dB)	−41.65	−50.31
Gain (dB)	7.66	8.32
HPBW (3dB)	78.9°	68.8°

Figure 10 illustrates the antenna radiation pattern with the metamaterial cover. As it can be observed in the Figure 10, the gain obtained is 8.32 dB at the frequency of resonance, the HPBW is 68.8 degrees and the main lobe orientation is zero.

Figure 11 indicates the radiation profile of the metamaterial-coated antenna in view of three dimensions. For comparing and better understanding, antenna outcomes in both types, with and without metamaterial, have been listed in Table 2. The beneficial impact of the usage of metamaterial cover in the function of the antenna is evident, according to Table 2.

As can be seen, gain has improved around 8.6 percent on the microscope patch antenna by supporting this designed structure. HPBW has also significantly decreased by about 12.8 percent. Such angle change has led to a significant improvement in the direction and a better centralization of the main beam of radiation pattern. In most of the high frequency applications, the more the antenna radiation can be transferred in the direction defined and intended or be oriented, the more the antenna would be preferable. Furthermore, in the frequency of resonance the S_{11} value also has a desirable change in this structure due to the metamaterial antenna configuration and indicates a suitable impedance matching [1].

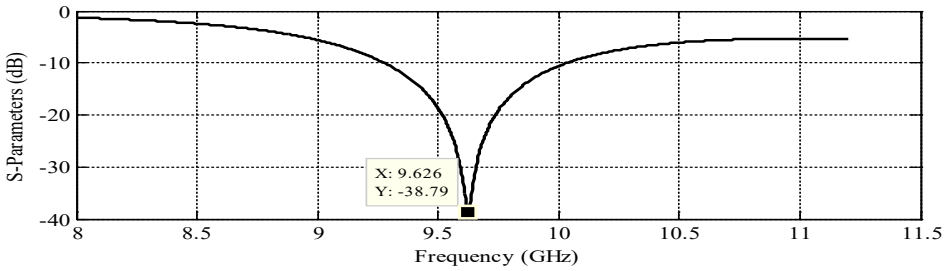


Figure 12: The return losses in the two-layer SRR mode

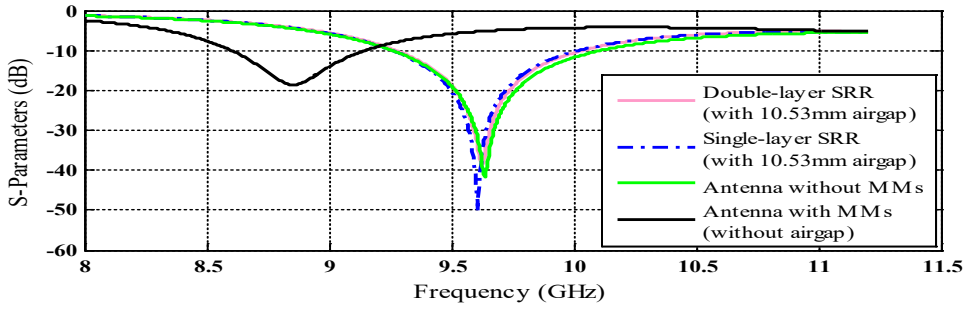


Figure 13: A comparison of the return losses and resonance frequencies of four mentioned antennas

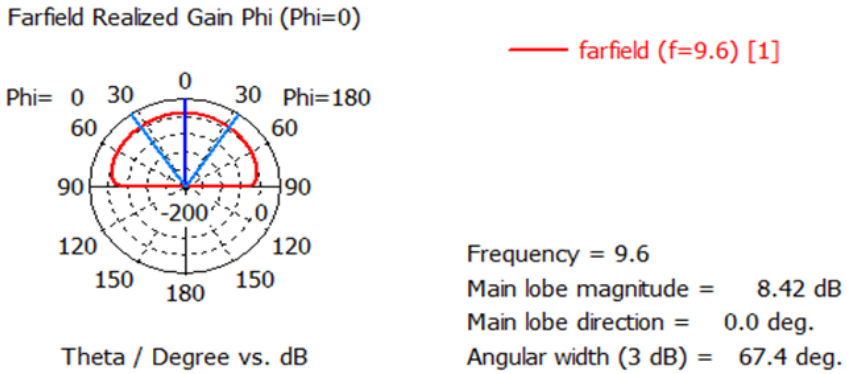


Figure 14: The pattern of radiation in double-layer SRR

3.2. Double-layer SRR effect

The metamaterial cover in this section comprises of two coats of SRR on the lower and upper plane of Teflon.

Because there are 15 SRR unit cells in each Teflon surface, the total number will be 30 unit cells. It has been positioned in the presumed airspace afterwards and outcomes have been compared to the monolayer SRR.

Figures 12 and 14 display the results of return losses and radiation pattern respectively. Also in order to comparing and gaining a better perception of the function of four mentioned types of antennas, results of the return losses have been represented in Figure 13.

The analogy of results between the single-layer SRR and the double-layer SRR can be seen in Table 3.

Table 3: Comparison of modeled SRR values in single-layer mode with double-layer mode

Parameters	One layer of SRR	Two layers of SRR
Return loss (dB)	−50.31	−38.78
Gain (dB)	8.32	8.42
HPBW (3dB)	68.8°	67.4°

As shown in Table 3, when the Teflon slab contains two upper and lower SRR layers, although the return loss has reached to -38, it would be in a highly admissible condition nevertheless, and the optimal output of the antenna is not harmed. The gain of antenna has had increasing pattern, plus HPBW has fallen by 1.4 degrees as well. Accordingly, in enhancing the radiation pattern, SRR in double-layer mode could become more efficient style [1]. In the current work, taking into account changes in antenna size and dielectric constant, the metamaterial-free antenna not only had stronger radiation than the compared article patch antenna in [19], but also some modifications in the form and quantity of the SRRs in the metamaterial antenna, as well as the amount of the air gap between the antenna and the cover, allowed noticeable improvements to be made. When we juxtapose the findings of metamaterial antennas in both Tables 1 and 3 (one-layer mode of SRR), we will find that the gain from present work has improved in comparison with the paper described. We decreased the HPBW by around 9.47 percent than the aforementioned article as well. Reducing HPBW implies that the centralization and power of the beam has improved [1]. For most of high-frequency antennas, that would be our key objective.

3.3. Effect of increasing the size and number of SRRs

At this stage, in order to modifying the previous operations, some changes have been applied to the shape, size, and number of unit cells. For this purpose, a piece of copper wire has been added between two squares. Also the number of unit cells has been increased from 5 * 3 unit to 5 * 4, and then to 5 * 5 unit, afterwards results have been investigated. New unit cell is shown in Figure 15. The results for the twenty and twenty-five resonators have been presented in the Table 4.

According to the Table 4, the radiation pattern has improved with the increase of size and number of SRR. But considering the value of S_{11} and comparing them in both twenty and 25 types, it is observed that as the number of resonators increases, the numerical value of S_{11} has changed. This means that return losses of antenna in metamaterial coating with 25 resonators have increased. Thus, it can be concluded that the increase in the number of resonators is faced with limitation, and if it exceeds a certain number, the antenna losses will increase.

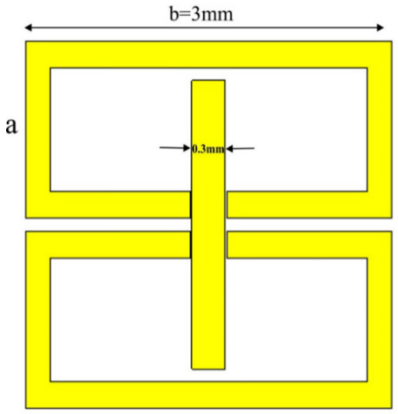


Figure 15: The modified unit cell by adding a piece of wire

Table 4: Comparison of the simulated results between metamaterial with 20 and 25 unit cells

Number of SRR	4*5 units	5*5 units
Frequency (GHz)	9.6448	9.6576
S_{11} (dB)	-23.66	-15.88
Gain (dB)	8.91	9.36
HPBW (deg)	60.2	55.9

But since in many space applications, antennas with sharp beamwidth are required for detecting and tracing purposes, it seems tempting to use this type of antennas as a lens. For this reason, it is not easy to ignore the sharpened beamwidth and high directivity, that is result of increasing the number of resonators. Therefore, to overcome this limitation, we have benefited a number of circular holes in the ground plane. The procedure is described in section 3.3.1.

3.3.1. Effect of inserting circular holes in the ground plane

The idea has been derived from the article [22]. The procedure of mentioned paper, has been to improve the antenna radiation pattern by using resonators along with creating nine holes in the ground plane. The use of circular holes, is an old method of limiting surface currents, which reduces the amount of return losses [22]. For this purpose, along with the use of metamaterial coating with 25 resonators, we have created nine circular holes with the radius of 1.5 mm on the ground plane. Figure 16 shows the metamaterial cover, and the holes created on the ground plane.

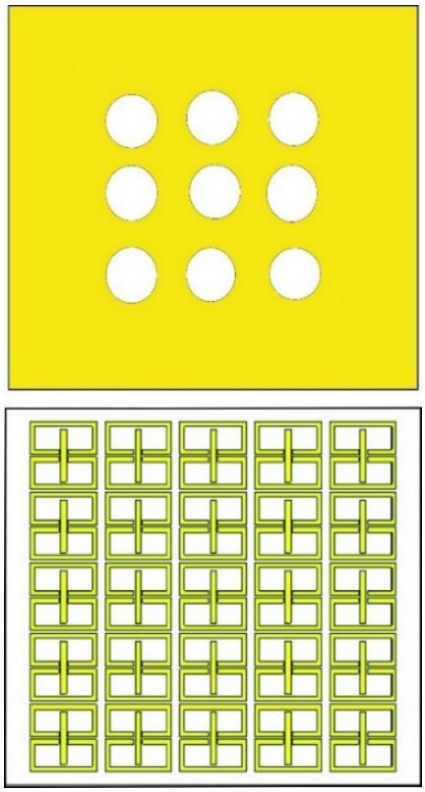


Figure 16: Circular holes created on the ground, and metamaterial coating with 25 unit cells

Table 5: Effect of different radii of holes on the return losses and resonance frequency

Radius (mm)	Frequency (GHz)	S_{11} (dB)
0.7	9.597	-15.74
0.8	9.635	-18.260
0.9	9.638	-19.94
1	9.635	-21.71
1.1	9.622	-24.29
1.2	9.603	-27.80
1.3	9.568	-29.24
1.4	9.539	-42.47
1.5	9.510	-38.46

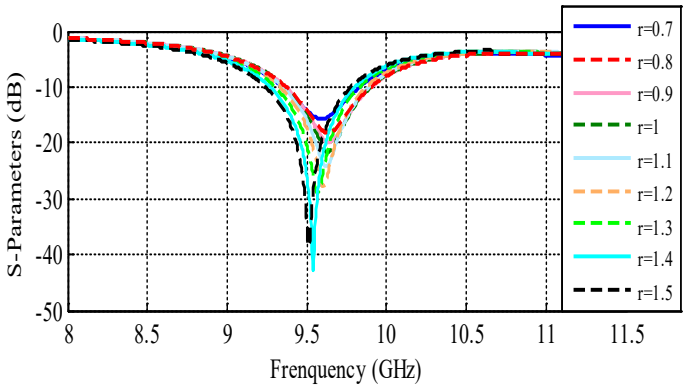


Figure 17: The return losses for different radii of circular holes

It should be noted that the radius of the holes can vary. The results of the effect of several different radii on return losses are presented in Table 5. Figure 17 also shows the return losses for the tabulated amounts derived from CST software.

As it is clear, the return losses have decreased dramatically as the holes' radii have increased. Frequency fluctuations are slightly visible as well. According to the Table 5, it can be understood that the hole creation significantly reduces the return losses and thus eliminates the antenna constraint caused by the increase of resonators.

It is worth noting that just because the frequency has also shifted slightly with the radius of the hole, in order to increase the accuracy of the radiation pattern results, we have extended the resonance frequency from 9.5 (for hole with radius of 1.5mm) to 9.6, which is the scope of this paper. This object can be easily achieved by altering the substrate dielectric constant from 2.2 to 2.1. For this purpose, we have replaced the Teflon of substrate with dielectric constant of 2.2 to PTFE existed in the software with dielectric constant of 2.1, in which case the frequency has reached 9.68. Also the amount of is about -32 which is quite acceptable.

Figure 18 demonstrates the radiation pattern of the antenna with the final improvement. Table 6 compares the results of this metamaterial antenna with those results in the [22].

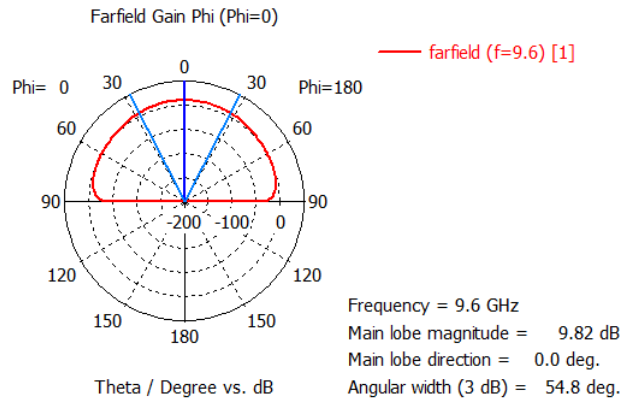


Figure 18: The radiation pattern of antenna including both 25 SRRs and 9 circular holes

Table 6: Comparison of simulated results between final improvement of this paper and paper [22]

antenna	Reference paper	Our work	Improvement (measured)
Return loss(dB)	-28	-32.78	-4.78 dB
Gain(dB)	9.6	9.82	0.22 dB
HPBW(deg)	68	54.8	19.5%

The creation of circular holes in the ground plane not only have significantly reduced the return losses, but in comparison with the Table 4 -when there is no hole- they have also increased the amount of antenna gain at a rate of 0.46 dB.

In general terms, the creation of the circular holes along with metamaterials consisting of large-sized and more-numbered resonators have impressively increased the efficiency of the antenna.

This method can therefore be regarded as a solution for reducing antenna losses and overcoming the antenna restriction due to the increase in resonator numbers.

3.4. Effect of Metamaterial layer on the patch antenna in hyperthermia application

Table 7 shows amounts of SAR obtained for different distances of Ag-hm. To better realization of the Table 7, results have been shown in Figure 19.

As it can be seen in the diagram, the lowest amount of output SAR is pertinent to 24 mm, namely maximum Ag-hm distance.

By reducing the distance between the metamaterial coating and the body tissue, the numerical value of SAR has increased relatively, so that at 6.25 mm, the amount of SAR has reached its maximum value of 15 w/kg.

It can be concluded from Table 7 that the distance between the tissue and the metamaterial antenna plays a significant role in the output SAR value. In other words, the SAR values required for different usages can be obtained by varying the distance.

In the next phase of the experiment, the hotspot of the lowest SAR amount obtained (5.2 w/kg), at the maximum distance between tissue and metamaterial, was estimated at approximately 4.8 w/kg. In the following, the rest of the SAR simulations values were normalized accordingly.

A few examples of the results in the normalized SAR value are given in Figure 20.

Table 7: SAR amounts obtained for different distances between metamaterial coating and body tissue

Air Distance [Ag-hm] (mm)	SAR Amount (w/kg)
24	5.2
20	9.25
18.75	8.99
16.66	10.1
15	10.7
12.5	11.25
10	12.6
8.34	13.9
6.25	15
5	14.5
1	10.1

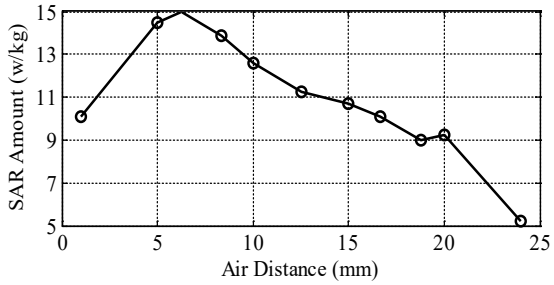


Figure 19: Schematic view of SAR amounts based on different gaps delineated with Table 7 data.

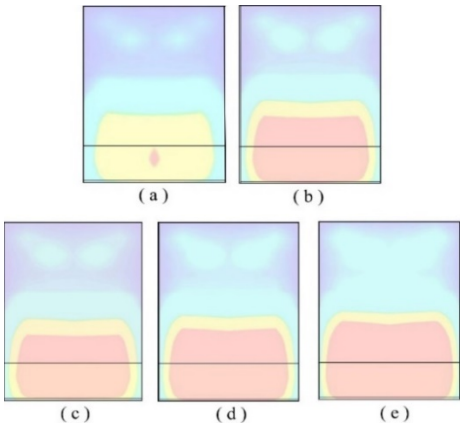


Figure 20: Normalized SAR based on Minimum SAR (in 4.8 w/kg). (a) Minimum SAR with value of 5.2 w/kg in distance of 24mm. (b) SAR amount: 9.25 w/kg in air gap of 20mm. (c) SAR amount: 8.99 w/kg in air gap of 18.75mm. (d) SAR amount: 10.7 w/kg in air gap of 15mm. (e) SAR amount: 13.9 w/kg in air gap of 8.34mm

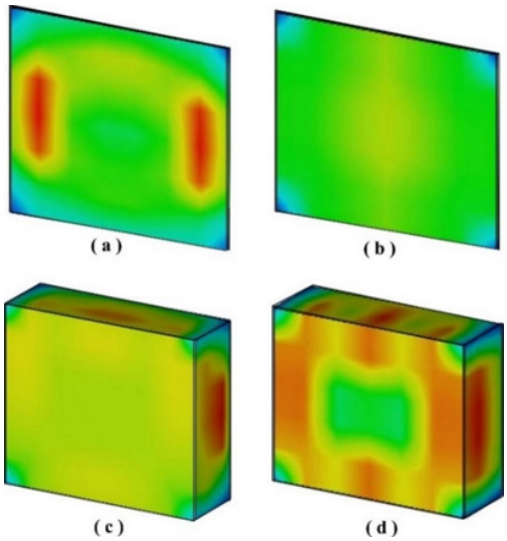


Figure 21: Comparison of heat permeation between usual and MMs antennas in two tissue layers of skin and fat. (a) Skin layer in usual simple antenna. (b) Skin layer in MMs antenna. (c) Fat layer in usual antenna (d) Fat layer in MMs antenna.

As it can be observed in Figure 20, with increasing the amount of output SAR, a larger area of tissue is also involved in heat. Therefore, with altering in this distance we will be able to increase or decrease the amount of the output SAR. To describe precisely, it will be feasible that the amount of heat required come under controlled to some extent.

The use of metamaterial coating on the antenna has another advantage, which is remarkable in its form, and that is the conduction of electromagnetic energy and the heat to the higher depths of the body tissue. This subject causes to reduce damages due to heat in long time to the surface tissues.
Figure 21 shows the two superficial layers of tissue, skin and fat, that compare the amount of transition heat between both types of simple microstrip antennas and metamaterial antennas.

As it is noticeable in Figure 21, in the usual antenna, the first surface layer of tissue namely skin has been burnt and the rest of heat has penetrated to the initial layers of fat, whereas in the metamaterial antenna, heat has penetrated into the deeper layers of fat and there is no burn on the skin tissue.

4. Conclusion

In this study, we took advantage of a type of metamaterial formation as a coating layer upon a microstrip patch antenna with coaxial feeding to upgrade the efficiency of antenna. The proffered small-size antenna was modeled in a resonance frequency of 9.6 GHz, whereby the metamaterial structure layer was used to achieve a highly desirable orientation. The gain of the antenna increased significantly as well, and the half-power beamwidth was decreased by more than 20 degrees than the normal patch antennas, which tend to be striking in spatial tracing applications considering the ordinary quantities of beamwidth in the usual antennas.
Using of metamaterials in antennas not only causes to reduce losses amount, increase the gain and sharpen the radiation as a lens in satellite usages, but also can alter the depth of penetration in medical applications, as well as acquire various SAR amounts, by adjusting the proper distance. In fact, more degrees of freedom will emerge for different applications. Specifically, by using different shapes of unit cells, other forms of these designs will present different outcomes.

References

[1] Jahangiri, Mahsa, and Saman Rajebi. "Effect of square srr structure on the radiation pattern of the patch antenna in the x-band." *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*. IEEE.
[2] VESELAGO, Viktor G. The electrodynamics of substances with simultaneously negative values of and μ. *Soviet physics uspekhi*, 1968, 10.4: 509.
[3] PENDRY, John B., et al. Magnetism from conductors and enhanced nonlinear phenomena. *IEEE transactions on microwave theory and techniques*, 1999, 47.11: 2075-2084.
[4] SMITH, David R., et al. Composite medium with simultaneously negative permeability and permittivity. *Physical review letters*, 2000, 84.18: 4184.
[5] ISLAM, Mohammad Tariqul, et al. A negative index metamaterial-inspired UWB antenna with an integration of complementary SRR and CLS unit cells for microwave imaging sensor applications. *Sensors*, 2015, 15.5: 11601-11627.
[6] ENGHETA, Nader. Metamaterials with negative permittivity and permeability: background, salient features, and new trends. In: *Microwave Symposium Digest, 2003 IEEE MTT-S International*. IEEE, 2003. p. 187-190.
[7] WITHAYACHUMNANKUL, Withawat; ABBOTT, Derek. Metamaterials in the terahertz regime. *IEEE Photonics Journal*, 2009, 1.2: 99-118.
[8] PATEL, B. D.; NARANG, Tanisha; JAIN, Shubhangi. Microstrip Patch Antenna-A Historical Perspective of the Development. In: *Conference on Advances in Communication and Control Systems*. 2013.
[9] CONSTANTINE, A. Balanis, et al. Antenna theory: analysis and design. *MICROSTRIP ANTENNAS, third edition, John wiley & sons*, 2005.
[10] RAHMAN, Md Mahfuzur. *Small size coupling feed and inductive shorting antenna for wide bandwidth, increased gain and efficiency with low specific absorption rate (SAR) operation*. 2016. PhD Thesis. Michigan Technological University.
[11] KARADE, Abha R.; ZADE, P. L. A miniaturized rectangular microstrip patch antenna using SSRR for WLAN applications. In: *Communications and Signal Processing (ICCSP), 2015 International Conference on*. IEEE, 2015. p. 1002-1004
[12] SMITH, D. R., et al. Electromagnetic parameter retrieval from inhomogeneous metamaterials. *Physical review E*, 2005, 71.3: 036617.
[13] HOSSAIN, M. I., et al. Low-SAR metamaterial-inspired printed monopole antenna. *Applied Physics A,* 2017, 123.1: 8.
[14] ALI, Tanweer; MOHAMMAD, Saadh AW; BIRADAR, Rajashekhar C. A novel metamaterial rectangular CSRR with pass band characteristics at 2.95 and 5.23 GHz. In: *Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2017 2nd IEEE International Conference on*. IEEE, 2017. p. 256-260.
[15] Nijhawan, Geeta, Siddharth Sagar Nijhawan, and Minu Sethi. "Hyperthermia Treatments." *Noble Metal-Metal Oxide Hybrid Nanoparticles*. Woodhead Publishing, 2019. 241-263
[16] Mendez, Hector Fabian Guarnizo, Mauricio Andrés Polochè Arango, and John Jairo Pantoja Acosta. "Hyperthermia Study in Breast Cancer Treatment." *Workshop on Engineering Applications*. Springer, Cham, 2018.
[17] Jaffar, Nabilah Abdul, et al. "An overview of metamaterials used in applicators in hyperthermia cancer treatment procedure." *2017 International Conference on Electrical, Electronics and System Engineering (ICEESE)*. IEEE, 2017.
[18] Ebrahimi-Ganjeh, Mohammad Ali, and Amir Reza Attari. "Study of water bolus effect on SAR penetration depth and effective field size for local hyperthermia." *progress in Electromagnetics Research* 4 (2008): 273-283.
[19] Paswan, Pradeep, et al. "Performance enhancement of coaxial feed microstrip patch antenna using left-handed metamaterial cover." *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science*. IEEE, 2014.
[20] SEO, Je Hyung, et al. Distribution of THz electric field in the split-ring resonator metamaterials based on the thin film geometry. *Current Applied Physics*, 2016, 16.3: 329-334.
[21] GARG, Ramesh. *Microstrip antenna design handbook*. Artech house, 2001.
[22] Fang, Chao, and Feng Xu. "Performance enhancement of coaxial feed microstrip patch antenna using left-handed material and PBG circular holes." *2014 IEEE International Conference on Communiction Problem-solving*. IEEE, 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 317-324 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Stress Response Index for Traumatic Childhood Experience Based on the Fusion of Hypothalamus Pituitary Adrenocorticol and Autonomic Nervous System Biomarkers

Noor Aimie Salleh*,[1], Malarvili Balakrishnan [1], Anna C. Whitttaker [2]

[1]*School of Biomedical Engineering and Health Science, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru Malaysia*

[2]*Faculty of Health Sciences and Sport, University of Stirling, Stirling, United Kingdom*

ARTICLEINFO

Article history:
Received: 30 September, 2019
Accepted: 19 January, 2020
Online: 07 February, 2020

Keywords:
Heart Rate Variability
Salivary Cortisol
Stress
Biomarkers
Fusion

ABSTRACT

Stress occurring in the early days of an individual was often assumed to cause several health consequences. A number of reports indicated that having to deal with unfavourable events or distress situation at a young age could tweak stress responses leading to a broad spectrum of poor mental and physical health condition. Therefore, changes identified within stress response were recommended to be taken as a measure in regulating and managing such health situation. This study combines the biomarker that represents both autonomic nervous system (ANS) and hypothalamic-pituitary-adrenocorticol (HPA) as a single measure to classify the stress response based on traumatic childhood experience and propose a stress response index as a future health indicator. Electrocardiograph (ECG), blood pressure, pulse rate and salivary cortisol (SCort) were collected from 12 participants who had traumatic childhood experience while the remaining 11 acted as the control group. The recording session was done during a Paced Auditory Serial Addition Test (PASAT). HRV was then computed from the ECG and the HRV features were extracted. Next, the best HRV features were selected using Genetic Algorithm (GA). Biomarkers such as BP, PR and SCort were then integrated with 12 HRV features picked from GA. The integrations were conducted using two fusion methods which are Euclidean distance and serial fusion. The differences in reaction of the fused features were then identified. Based on the result, the Euclidean distance (ed) which is the fused feature by the parallel fusion, displayed the most efficient reaction with accuracy, sensitivity, and specificity at 80.0%, 83.3% and 78.3%, respectively. Support Vector Machine (SVM) was utilized to attain such result. The fused feature performance was then fed into SVM which produced indexes on stress responses. The result retrieved from these indexes acts as a measure in handling future health deliverability and perhaps could eventually enhance the health care platform for midlife individuals.

1. Introduction

This paper is an extension of the work [1] originally presented in the IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2018. Stress is identified as a biological and mental reaction to any threat or external pressure. During normal conditions, stress encourages a person to concentrate, stay active and attentive. Nevertheless, in critical events, stress assists in powering up an individual and this may lead to the ability in saving someone's life. However, stress becomes damaging to the physical and psychological health of an individual when the body fails to cope and solve problems [2]. Stress can initially be a positive contributing factor in an individual's life but due to immoderate demands that complies with the advanced and modern world eventually becomes a major threat. Many acute diseases have been listed as the effect of stress. This includes cardiovascular diseases, diabetes, asthma, anxiety disorders and depression [3].

Stress is classified into three types which are known as acute, episodic and chronic [3,4,5]. Acute stress is known as stress occurring at a short period of time. Factors contributing to this kind of stress are usually work-related, athletic performances and examination. Episodic stress occurs due to repetitive acute stress, like that of a daily work stress. Chronic stress is related to a high intensity stress factors that develops over a long period of time. This

*Corresponding Author: Noor Aimie Salleh Email: aimie@biomedical.utm.my

can be damaging to the physical and psychological well-being of an individual [6]. Factors that can lead to development of chronic stress are marital affairs, work stress, financial condition [7], and traumatic life experiences [8, 9] such as an unpleasant childhood incident.

Several researchers have been caught up studying stress factor developed from traumatic childhood experiences. Recent studies in illness management and prevention indicates that being stressed at an early age may lead to a series of health issues [10, 11, 12, 13], and alterations in physiological behaviour [14, 15, 16, 17]. This fact is significant in developing methods that could improve healthcare management for individuals approaching adulthood, since there is an indicator of the current health status of the particular child. Therefore, this study focuses on traumatic childhood experience which is classified as chronic stress. In sequence to investigate the stress response of individual with traumatic childhood experience, mental stress test which is an acute stress is used.

Early life stress has shown a number of health outcomes to be predictive [18, 11]. Studies conducted a few years ago showed that individuals with unfavourable and stressful situation tend to fall in a broad spectrum of mental and physical health problems which eventually lead to alcohol abuse, obesity, depression, smoking, substance abuse, intrusive sexual acts [19], as well as acute illnesses like cardiovascular disease [20, 21]. The output of these studies has brought latest discovery that indicate children who are vulnerable to potentially traumatic events may experience cardiovascular stress reactivity due to acute physiological processes. There are a number of cases that displayed changes in cortisol as well as severe mental stress that increases heart rate within individuals who had at least one or more traumatic incident [14, 15]. Nevertheless, studies also showed changes in cardiovascular reactivity to stress task amongst children who were exposed to aggressive adolescent behaviours [17]. These evidences indicate that stress reactivity is detected via incidents taking place at an individual's tender age, and that the immensity of stress reactivity is in relation to stress exposure and may influence health of an individual in years to come [22, 23]. Since there are compelling evidences on this matter, detecting changes in stress response due to traumatic events of a child is significant to be utilized as an indicator for better healthcare.

It is important to pick the most suitable biomarker to identify and determine changes taking place in stress response. Heart rate variability (HRV) is one of the biomarkers proven to be the most powerful and depicts autonomic nervous system [24-27]. However, there is not much research on the application of HRV in measuring stress responses among children with traumatic incidents. In fact, only conventional frequency domain method was applied [28, 29]. Hence, both conventional and improved approaches extracted from HRV analysis should be applied in measuring stress responses of a child who had been through traumatic incidents.

Integrating two of the major body systems known as autonomic nervous system (ANS) and hypothalamic-pituitary-adrenocorticol (HPA) is strongly recommended as it plays an important role in regulating stress response [30,31]. Although numerous studies indicate that in measuring stress responses, multiple biomarkers were used. In conjunction to that, the results for each biomarker were analysed singly [32, 33].

Therefore, this study proposes a latest fusion method for stress response classification on traumatic childhood occurrence. Since children's traumatic incident causes inconsistency in stress response, the health status during adulthood can be determined via a proposed stress response index which could indicate the future health condition and lead to preventative measurements as well as further diagnosis that may refine the healthcare management. The objective of this study is to integrate both HPA and ANS so as to classify stress responses among children who faced traumatic experiences along with a stress response index that would be able to indicate health statuses in the later years of these individuals.

2. Method

2.1. Participants

609 students were screened with lifetime adversity inventory based on C-DIS-IV items and Childhood Traumatic Questionnaire (CTQ) [34]. This group of participants is made up of 511 females and 98 males (age, M = 19.2, SD = 2.99 years). Participants who scored within 2 to 5 and 2 or higher on the lifetime adversity measure and CTQ, respectively, were identified as individuals with traumatic childhood incidents whereas those who scored 0 on the lifetime adversity measure and CTQ were made the control group [14, 34]. A total of 23 participants, 12 participants who had traumatic childhood experience and 11 participants who were from the healthy control group. These participants have no history of cardiovascular illnesses, incurable diseases or other acute infection, and a current endocrine or immune disorder. Participants' consent was obtained and the study was approved by the University of Birmingham Ethnics Committee.

2.2. Mental Stress Test

The psychological stress test in this research was conducted with the Paced Auditory Serial Addition Test (PASAT) and it usually takes about 10 minutes to complete. Participants were given a series of single-digit numbers and were requested to add the digit to the numbers they previously heard. They were to say their answers out loud.

2.3. Procedure

A standard three-lead ECG placement and blood pressure (BP) cuff, were attached to the participants in sitting position. The test began with a 10 minutes adaptation period followed by another 10 minutes of resting baseline phase. Three saliva samples were collected from every subject with the first sample obtained at the last minute of the baseline phase. 10-minute PASAT was then completed and saliva sample was obtained once again. The recovery phase took place for the next ten minutes followed by last saliva sample collection. Participants' BP and heart rate (HR) were recorded every two minutes, four times for each phase, whereas ECG readings were recorded throughout the test.

2.4. Saliva Sampling and Cortisol Assays

In order to obtain samples of saliva, participants had salivette dental swab positioned in their mouth which was gently chewed for one minute and placed in the salivette tube. Three stimulated saliva samples were obtained in this study whereby first sample via a 10-minute baseline phase. As for the second and third sample, it was obtained right after stress test and initially 10 minutes to task, respectively. The salivettes were then centrifuged for five minutes at room temperature and 3500 rpm before being aliquoted. It was then made to freeze at -20 degrees until the time to be assayed.

The assayed cortisol was then duplicated by an enzyme-linked immunosorbent assay (ELISA) via a commercial kit (DRG Diagnostics).

2.5. Pre-Processing: From ECG to HRV

HRV were being quantified via MATLAB software based on obtained and pre-processed ECG signals. A notch filter was used in eliminating the 50 Hz power line interference. Pan and Tompkin's algorithm were applied to identify the QRS waves database [35].

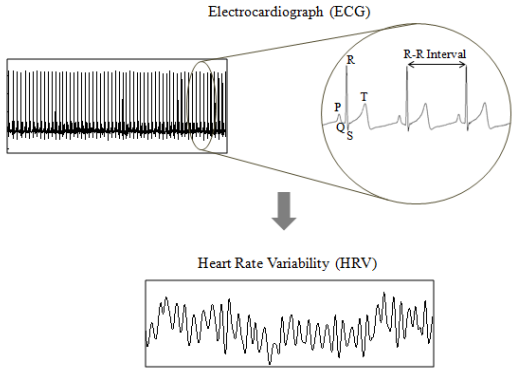


Figure 1: Derivation of Heart Rate variability for Electrocradiograph

2.6. Feature Extraction

In this research, HRV feature extraction is carried out through time, frequency, non-linear time-frequency and Wavelet analysis.

2.6.1. Time-Domain Analysis

The time-domain analysis was measured with Standard Deviation of the Normal-to-Normal intervals (SDNN), Standard Deviation of the Average of Normal-to-Normal intervals (SDANN), Root Mean Square Successive Difference (RMSSD) and HRV triangular index (HTI) [26] and computed via recommendation by Task Force (1996).

2.6.2. Frequency Domain Analysis

For the frequency domain analysis Autoregressive (AR) spectral analysis was applied [36,37]. AR model is interpreted as follow:

$$x[n] = \sum_{i=1}^{N} a_i x[n-1] + \varepsilon[n] \quad (1)$$

where, *x[n]* is the current value of the HRV time series, *a1, . . . , aN* are the predictor coefficients, *N* is the model order. With this method, the power spectral density (PSD) can be presented at a precise estimation. This would allow the feature extraction process to be carried out at ease as shown in Figure 2 [38].

Next, nine features were obtained via PSD extraction. These nine features are VLF (0.003-0.04 Hz), LF (0.040-0.15 Hz), HF (0.15-0.4 Hz), the standard unit (n.u.) of LF and HF (LFnu and HFnu), LF/HF ratio, Total Power (TP) [26, 38,39], log LF (lnLF) and log HF (lnHF)[40].

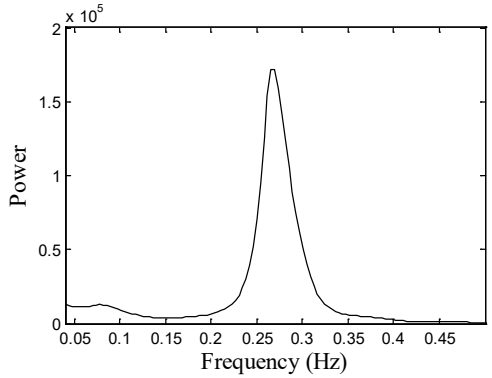


Figure 2: Power Spectral Density (PSD) using AR

2.6.3. Time-Frequency Domain Analysis

Since HRV contains slow varying signal, Modified B-Distribution (MBD) tend to be one of the most efficient Time Frequency Domains (TFDs) in exhibiting a high time-frequency resolution [41]. MBD's fundamental is [41]:

$$g(\nu,\tau) = \Gamma(\beta + j\pi\nu)^2 / \Gamma^2(\beta) \quad (2)$$

where, gamma function is identified as $\Gamma$and β is a positive real number between 0 and 1 that monitors the commutation between components' resolution and cross-term suppression [41].

Figure 3 shows an example of TF plotted via application of MBD to obtain HRV signals that were recorded during resting baseline. TFD measured via TFD-based Shannon and Renyi entropy led to an overall computation of TFD and as well as entropy within LF and HF domain [42,43].

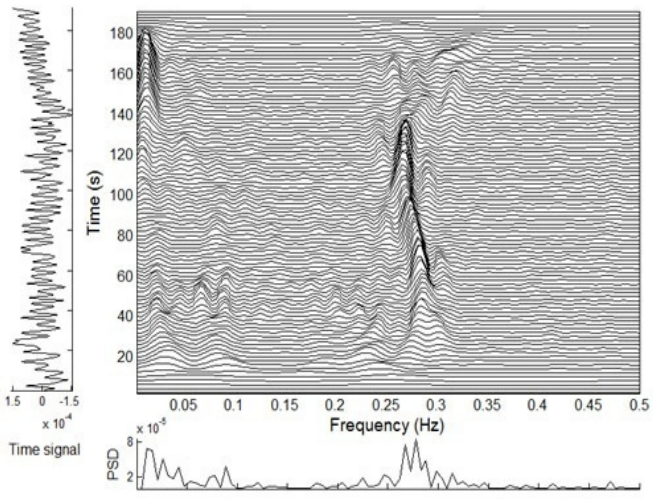


Figure 3: Time and frequency details via HRV's time-frequency distribution

2.6.4. Wavelet Transform Analysis

Wavelet analysis bring forth the localizations of time and frequency resulting in wavelet coefficients which can be utilized as features in classifiers [44,45]. HRV signal can be disintegrated using wavelet family $\psi a,b$, a basis function that involves dilatation and translations of a distinctive and valid mother wavelet $\psi(t)$. This process is defined as,

$$\psi_{a,b} = \frac{1}{\sqrt{a}}\psi\frac{t-b}{a} \quad (3)$$

where, scale and location is indicated by *a* and *b*, respectively [44, 45]. Basis function of DWT is represented at scale 2-m whereas, n indicates the time instant specified as follow,

$$\psi_{m,n} = \int_{-\infty}^{\infty} 2^{-\frac{m}{2}}\psi(2^{-m}t - n) \quad (4)$$

Hence, DWT's signal x(t) is specified as follow:

$$T_{m,n} = \int_{-\infty}^{\infty} x(t)\psi_{m,n}(t)dt \quad (5)$$

A discrete wavelet was transformed from eight-order Daubechies mother wavelet (db8) to five levels before being implemented to HRV. This transformation causes signal disintegration of five-level wavelet. Following this, using the suggested HRV frequency bands, Detail (D) and Approximate (A) coefficients were reassembled [26]. The A5 coefficient allowed rebuild of VLF range at < 0·04 Hz; D5 and D4 coefficients for LF range at 0·04–0·15 Hz; D23, D2, and D1 coefficients for HF range [46]. Hence, a total of 83 features were obtained from the HRV signal.

2.7. Feature Selection

The feature selection process is carried out via Genetic Algorithm (GA) MATLAB Toolbox. Iteration process occurred whereby chromosome is being engineered for production of new population via application of genetic functions like crossovers and mutations. Theoretically, healthy species will make through the evolution process whereas the weak ones disintegrate.

An initial population of chromosomes was randomly selected in order for GA's process to begin. Encoded bit strings usually depicts chromosome that is made up of sequence of features. These features each is classified as genes. In each chromosome, based on positional index, a gene value of '1' determines that the feature selected is positioned at '1' but the feature is not picked if the gene value is found to be at '0'. When iteration process takes place, combinatorial set of genes or features in the current population, identified as individual was observed via fitness function.

For the current work, the fitness function (*FitFunc*) was computed using kNN-based classification error with $k = 3$ [47,48,49] which is defined as:

$$FitFunc = \frac{\alpha}{N_f} + \exp\left(-\frac{1}{N_f}\right) \quad (6)$$

where, kNN-based classification error is indicated by α and Nf is the number of elements in selected features. In record, one study proposed the application of SVM classification error in extracting fitness function [50]. Unfortunately, despite being a substantial machine learning tool, SVM needs a durable computational load. Hence, in this research, this method was not applicable due to the constraint and fact that GA consumes computational space.

The initial populations of chromosomes were then ranked upon creation. The fittest individuals known as elitism solutions were selected from these ranks for survival in the next generation. Individuals that were not selected formed crossover and mutation solutions by genetically going through functional crossover and mutation. New generations were then formed by individuals selected through the elite, crossover and mutations.

GA operator's optimization was carried out to detect the prime value of initial population, crossover and mutation. Table 1 indicates the tested values extracted via references from studies carried out previously [47, 49, 51]. The outcomes of chosen features integrating with tested parameters were recorded and chromosomes with the optimum outcome were utilized in the next stage.

Table 1: Tested parameter figure utilized selection of HRV feature in GA.

Parameter	Tested Value
Initial Population	20
	30
	40
	50
	60
Crossover	0.1
	0.01
	0.001
Mutation	0.8
	0.9

Selection tool was required in choosing the best individual. Hence, in this research, two individuals or chromosomes were picked as defeater of the tournament by applying tournament selection of size 2. This selection was applied because of its clarity, rapidity and competency [47, 48]. The process is carried out in repetition until formation of new population is achieved.

Individuals like elite, crossover, and mutation kids were added to create the new population.

New Generation= Number (Elite kids) + Number (Crossover kids)…. + Number (Mutation kids) (7)

New Generation Score = Fitness (New Generation) (8)

This process continued until GA achieved the condition that puts it to stop. The stopping conditions utilized are Maximum Number of Generations (MaxNG) and Stall Generation Limit (SGLim). Both were set at 100 and 10, respectively [49]. If the mean alteration in the fitness gain between the chromosomes over SGLim generations was below than or equivalent to 0.000001 which is the value of Tolerances Function (TolFun), then the GA will be dismissed. The process ends if TolFun also known at termination tolerance, has a function gain whereby if $|f(xi) - f(xi+1)| <$ TolFun. Since GA observes the variance in values of fitness within all generations. The mean of these variances for 10 generations were recorded. If the recorded values were less than or equal to 0.000001, GA will be dismissed. This indicates that

resemblance in fitness value also known as genetic homogeneity among chromosomes of the generation have the most excellent chromosome and therefore coincides with GA.

2.8. Feature Fusion

In obtaining reactivity demarcation, the features were measure based on normalization equation as follow [52, 53]:

$$Y_n = \frac{Y_{stress} - \hat{Y}_{rest}}{\hat{s}_{rest}} \quad (9)$$

where, sample feature vector throughout stress texting is Y, the average of sample feature during resting baseline is indicated as $\hat{Y}$, and standard deviation of feature vector throughout resting baseline is indicated as $\hat{s}$.

Linking feature vectors of the biomarkers obtained from HRV feature extraction (h), and BP, PR and SCort (b, p and c, respectively) led to construction of serial fusion, whereas parallel fusion was designed by integrating feature vectors and complex vector. To introduce combined features with complex vector, the following equation, c = a + *i*b, where a and b are two varied feature vectors of the same sample S and *i* is an imaginary unit, was applied.

The purpose of fusion was to integrate the essentials within ANS and HPA leading to a finer presentation of the elements. Hence, the differences are obtained via fusion of biomarkers. ANS biomarkers are HRV, BP and PR, whereas HPA biomarker is cortisol.

2.9. Classification

In order to speculate an input's output group with existing data, a classifier is required. In this research, SVM classifier was used to classify stress response between two types of group which are of individuals with traumatic childhood incidents and a control group. Other classifiers like Naive Bayes [54, 55] and k-Nearest Neighbour (kNN) [56, 57] were used in this study to distinguish and prove that SVM is the most effective classifier. The evaluations of classifier were carried out by applying 10-fold cross validation [46].

2.10. Stress Response Index

The stress response index is depicted via classifiers that indicate greater deliverability. The future health is then predicted via irregularity patterns in stress response that were detected within the index.

3. Result and Discussion

3.1. HRV Feature Extraction

From the HRV feature extraction, a total of 83 features that has been extracted. Previous researches indicated that most features selected in obtaining HRV classifications were of conventional methods which are based on time and frequency analysis [58-60]. Nevertheless, there were studies that indicate the application of feature analysis that are more advanced such as that of wavelet, time frequency, and non-linear [46, 61, 62]. However, in this study, obtaining HRV features were carried out with the fusion of both conventional and newest advance analysis.

3.2. HRV Feature Selection

From genetic algorithm, 12 features were selected which are RMSSD from time domain analysis, Normalized unit of high frequency (Hfnu) from AR PSD, Mean LF, SDNN D1, SDNN D4, Apen (Approximate entropy) D4, SampEn (sample entropy) D3, Kurtosis D4, SkewD5, from wavelet and ShanEn LF, ShanEn (Shannon entropy) HF, ShanEn LFHFr from TFD.

3.3. Feature Fusion

The extracted HRV features via GA feature selection were then made to integrate with biomarker obtained in this study to produce stress response index. The initial biomarker fusion was carried out with the application of Euclidean distance (ed) classified as parallel fusion followed by serial fusion, a regular and simple fusion method, carried out for resemblance identification. Theoretically, the fusion of biomarkers that depicts ANS and HPA was applied in this research. This is because both ANS and HPA are biological systems with their own pathological characteristics and are strongly connected. Hence, a booming stress response index was likely to be extracted from the fusion of these systems [30, 31, 63]. Biomarkers depicting ANS were made to integrate with biomarker that depicts HPA, known as, Scort. With this fusion, HRV-Scort, BP-Scort and PR-Scort, were produced.

Table 2 shows the result of integrated biomarkers via the application of SVM in the context of accuracy, sensitivity and specificity. Outstanding performance can be seen with application of Euclidean distance by HRV-Cort as accuracy, sensitivity and specificity scored 80.0%, 83/3%, and 78.3%, respectively. The table also indicated that, biomarkers like PR, BP and HRV representing ANS with Scort had poor performance. From this result, it can be seen that HRV is a strong biomarker and can be utilized for ANS.

Following this, a single biomarker was used to measure the differences between outcomes of new fused feature and stress response classifications. Results indicated that the latter had better performance than single biomarker. However, since SVM model indicated excellent outcomes in categorizing fused biomarker (HRV-SCort), it was then utilized in exhibiting stress response index.

Table 2: Outcomes of SVM classifiers applied in fused biomarker

Fusion Method		Acc %	Sen%	Spe%
Serial	HRV-Cort	63.3	51.7	75.0
	BP-Cort	67.2	68.3	66.7
	PR-Cort	58.3	66.7	48.3
	HRV-BP-PR-Cort	65.0	58.3	71.7
Euclidean distance	HRV-Cort	80.0	83.3	78.3
	BP-Cort	21.7	13.3	28.3
	PR-Cort	59.4	91.7	26.7
	HRV-BP-PR-Cort	68.9	81.7	60.0

*Notes: Acc: accuracy, Sen: sensitivity, Spe: Specificity

3.4. Support Vector Machine Model Fusion

The fused feature vector was then utilized as model with SVM MATLAB toolbox application. The input was placed in SVM model. Values of alpha (α), support vector, scale factor, shift and bias were obtained. The stress response index was recorded with

SVM classification (C_{SV}) algorithm as reference. The algorithm is interpreted as follow:

$$C_{sv} = \sum_i \alpha_i k_l(s_i, x) + b \quad (10)$$

where, si acts as support vector, α is the support vector's weight, b signifying bias and defined as the interception of hyperplane that forms normalized data space by division of two groups, and x is the recorded scale training vector from

$$(\boldsymbol{f}_e)_{scaled} = scale\ factor\ \times (\boldsymbol{f}_{Tr} + shift) \quad (11)$$

Meanwhile k1 is the dot product which is also recognized as linear kernel, as follow:

$$\text{Linear kernel} = \text{dot product} = \text{conj}(s_i).*x = s_iT.*x \quad (12)$$

Hence, based on (13), proposed stress response index (SRi) was indicated as:

$$\text{SR}_i = \sum_i \alpha_i \times (s_i{}^T .* [(\boldsymbol{H}_R) + i(\boldsymbol{C}_R)]_{scaled}) - 0.057 \quad (13)$$

where HRV reactivity occurred via HR and SCort reactivity is via CR.

Figure 4 shows the suggested Stress Response Index (*SRi*) of HRV and salivary cortisol (HRV-SCort) with *ed*-SVM of children's traumatic incidents. The figure indicated that two groups of stress response were divided by index boundary at a value of '0'. Normal stress response was identified with index sample ≥ 0, while other samples exceeding normal stress response were categorized as irregular stress response. Irregular stress response is often made up of participants who had no traumatic experiences during their early years.

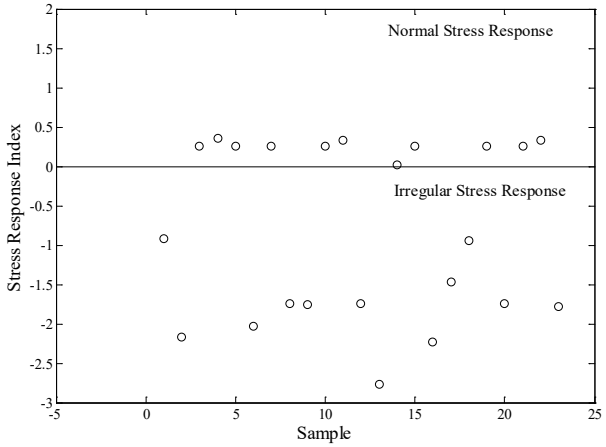


Figure 4: *SRi* of HRV and HRV-SCort with ed-SVM of traumatic childhood experience

A recent stress response index in reference to traumatic childhood incidence was proposed with the obtained comprehensive feature extraction. This extraction is made up of robust HRV feature with the application of GA and integration of *ed* and SVM. Linear regression was referred in designing the stress response index. Hence, the scalar value is inconsistent and changes based on sample's size. Assumption made is that the greater number of sample leads to stress response index being more accurate.

Nevertheless, alcohol intake, smoking, diet, present stress and psychosocial situation may influence the precision of stress response index [64]. This is a take on for researchers to further study on the factors influencing stress response as well as conducting the study with a bigger sample size to create a more complete index that acts as the fundamentals of stress response index. Furthermore, the reaction of traumatic childhood incidents within an individual against the stimulated stress requires further comprehensive research. Chances are that these individuals could be more at risk physically and mentally or they are able to manage themselves being in such situation. Findings from this research can be utilized to upgrade the approach within adult's healthcare [65].

4. Conclusion

This research proposed stress response index based on the combination of biomarker representing both autonomic nervous system and hypothalamus pituitary adrenocortical to classify stress response for traumatic childhood experience. 12 HRV features, form time, frequency, Time, frequency, TF and wavelet were extracted then combined with other biomarkers, blood pressure, pulse rate, and salivary cortisol. The biomarker combination represents two main body systems which are autonomic nervous system and hypothalamus pituitary adrenocortical axis. The classification was performed using support vector machine and the result proved that the combination between HRV and Salivary Cortisol demonstrated the highest performance with 80.0% accuracy, 83.3% sensitivity and 78.3% specificity. Finally, The fused vector was fed into the SVM model to develop the stress response index. This index could be used to predict the future health and therefore allow the individual to take more cautious about his/her health.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was funded by the Ministry of Higher Education, Malaysian Government (FRGS: R.J130000.7745.4F943) and funding for research from School of Sport, Exercise & Rehabilitation Sciences, University of Birmingham, UK.

References

[1] Aimie-Salleh, Noor & Balakrishnan, Malarvili & Whitttaker, Anna, "Stress response index for adverse childhood experience based on fusion of biomarkers". IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 140-145, 2018. DOI: 10.1109/IECBES.2018.8626657.

[2] T. D. Washington, Psychological Stress and Anxiety in Middle to Late Childhood and Early Adolescence: Manifestations and Management. Journal of Pediatric Nursing, 24(4), 302-313, 2009.

[3] N. Schneiderman, G. Ironson, and S. D. Siegel, Stress and Health: Psychological, Behavioral, and Biological Determinant. Annual Review of Clinical Psychology, 1, 607–628, 2005.

[4] A. Fernandes, R. Helawar, R. Lokesh, T. Tari, and A. V. Shahapurkar, A Survey on Analysis of Stress. International Journal of Latest Trends in Engineering and Technology, Special Issue, 31-35, 2013.

[5] V. H. Rice, Theories of Stress and Its Relationship to Health. In Rice, V. H. (Ed.) Handbook of Stress, Coping, and Health: Implications for Nursing Research, Theory, and Practice, Detroit: SAGE Publications, Inc, 23-42, 2012.

[6] A. Baum, M. K. O'Keeffe, and L. M. Davidson, Acute Stressors and Chronic Response: The Case of Traumatic Stress. Journal of Applied Social Psychology, 20(20), 1623-1724, 1990.
[7] M. R. Salleh, Life Event, Stress and Illness. The Malaysian Journal of Medical Science: MJMS, Pmc, 15(4), 9-18, 2008.
[8] R. Finlay-Jones, and G. W. Brown, Types of Stressful Life Event and the Onset of Anxiety and Depressive Disorders. Psychol Med, 11(4), 803-15, 1981.
[9] G. Andrews, C. Tennant, D. M. Hewson, and G. E. Vaillant, Life Event Stress, Social Support, Coping Style, and Risk of Psychological Impairment. The Journal of Nervous and Mental Disease, 166(5), 307-16, 1978.
[10] R. F. Anda, D. W. Brown, S. R. Dube, J. D. Bremner, V. J. Felitti, and W. H. Giles, Adverse Childhood Experiences and Chronic Obstructive Pulmonary Disease in Adults. American Journal of Preventive Medicine, 34(5), 396-403, 2008.
[11] F. V. J. Felitti, et al., "Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults: The Adverse Childhood Experiences (ACE) Study," American Journal of Preventive Medicine, 14, 245-258, 1998.
[12] V. J. Felitti, Long-Term Medical Consequences of Incest, Rape, and Molestation. Southern MedicalJjournal, PubMed 84(3), 328-331, 1991.
[13] V. J. Felitti, Childhood Sexual Abuse, Depression, and Family Dysfunction in Adult Obese Patients: A Case Control Study. Southern Medical Journal, PubMed, 86(7), 732-736, 1993.
[14] W. R. Lovallo, N. H. Farag, K. H. Sorocco, A. J. Cohoon, and A. S. Vincent, "Lifetime adversity leads to blunted stress axis reactivity: studies from the Oklahoma Family Health Patterns Project," Biological Psychiatry, 71, 344-349, 2012.
[15] W. R. Lovallo, "Early life adversity reduces stress reactivity and enhances impulsive behavior: Implications for health behaviors," International Journal of Psychophysiology, 90, 8-16, 2013.
[16] L. J. Luecken, Childhood Attachment and Loss Experiences Affect Adult Cardiovascular and Cortisol Function. Psychosomatic Medicine, 60(6), 765-72. 1998.
[17] L. J. Luecken and D. S. Roubinov, "Hostile behavior links negative childhood family relationships to heart rate reactivity and recovery in young adulthood," International Journal of Psychophysiology, 84, 172-179, 02/11 2012.
[18] R. Anda, D. Williamson, D. Jones, C. Macera, E. Eaker, A.Glassman, and J.Marks, "Depressed Affect, Hopelessness, and the Risk of Ischemic Heart Disease in a Cohort of U.S. Adults," Epidemiology, 4, 285-294, 1993.
[19] R. F. Anda, et al., "The enduring effects of abuse and related adverse experiences in childhood," European Archives of Psychiatry and Clinical Neuroscience, 256, 174-186, 2006
[20] S. V. Batten, D. W. Brown, S. R. Dube, J. D. Bremner, V. J. Felitti, and W. H. Giles, "Childhood Maltreatment as a Risk Factor for Adult Cardiovascular Disease and Depression," Journal of Clinical Psychiatry, 65, 249-54, Feb 2004.
[21] M. Dong, et al., "Insights into causal pathways for ischemic heart disease: adverse childhood experiences study," Circulation, 110, 1761-6, Sep 28 2004.
[22] W. T. Boyce and B. J. Ellis, "Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity," Development and Psychopathology, 17, 271-301, Spring 2005.
[23] J. Ellis, et al., "Biological sensitivity to context: II. Empirical explorations of an evolutionary-developmental theory," Development and Psychopathology, 17, 303-28, Spring 2005.
[24] U. R. Acharya, et al., "Heart Rate Variability: A Review," Medical and Biological Engineering and Computing, 44, 1031-1051, 2006.
[25] A. P. Allen, et al., "Biological and Psychological Markers of Stress in Humans: Focus on the Trier Social Stress Test," Neuroscience and Biobehavioral Reviews 38, 94–124, 2014.
[26] Task Force, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology," Circulation, 93, 1043-1065, 1996.
[27] J. F. Thayer, et al., "A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health," Neuroscience & Biobehavioral Reviews, 36, 747-756, 2012.
[28] J. E. Carroll, et al., "Childhood abuse, parental warmth, and adult multisystem biological risk in the Coronary Artery Risk Development in Young Adults study," Proceedings of the National Academy of Sciences, 110, 17149-17153, 2013.
[29] K. Winzeler, et al., "Adverse childhood experiences and autonomic regulation in response to acute stress: the role of the sympathetic and parasympathetic nervous systems," Anxiety, Stress, & Coping, 30, 145-154, 2017/03/04 2017.
[30] N. Michels, et al., "Relation between salivary cortisol as stress biomarker and dietary pattern in children," Psychoneuroendocrinology, 38, 1512-1520, 2013.
[31] N. Michels, et al., "Children's heart rate variability as stress indicator: Association with reported stress and cortisol," Biological Psychology, 94, 433-440, 2013.
[32] A. T. Ginty, et al., "Disordered eating behaviour is associated with blunted cortisol and cardiovascular reactions to acute psychological stress," Psychoneuroendocrinology, 37,715-724, 5// 2012.
[33] A. C. Phillips, et al., "Cardiovascular and cortisol reactions to acute psychological stress and adiposity: cross-sectional and prospective associations in the Dutch Famine Birth Cohort Study," Psychosomatic Medicine, 74, 699-710, Sep 2012.
[34] D. P. Bernstein, et al., "Development and validation of a brief screening version of the Childhood Trauma Questionnaire," Child Abuse & Neglect, 27, 169-90, Feb 2003.
[35] J. Pan and J. W. Tompkins, "A Real-Time QRS Detection Algorithm," IEEE Transactions on Biomedical Engineering, 32, 230-236, 1985.
[36] U. R. Acharya, et al., "Automatic Identification of Cardiac Health Using Modeling Techniques: A Comparative Study," Information Sciences, 178, 4571-4582, 2008.
[37] R. H. Takalo and H. H. Ihalainen, " Tutorial on Univariate Autoregressive Spectral Analysis Export," The Journal of Clinical Monitoring and Computing, 20, 379-379, 2006.
[38] M. N. Castro, et al., "Heart rate variability response to mental arithmetic stress in patients with schizophrenia: Autonomic response to stress in schizophrenia," Schizophrenia Research, 99, 294-303, 2008.
[39] P. Karthikeyan, et al., "Analysis of Stroop Color Word Test-Based Human Stress Detection using Electrocardiography and Heart Rate Variability Signals," Arabian Journal for Science and Engineering, 39, 1835-1847, 2014.
[40] J. A. J. Heathers, "Smartphone-enabled pulse rate variability: An alternative methodology for the collection of heart rate variability in psychophysiological research," International Journal of Psychophysiology, 89, 297-304, 2013.
[41] B. Barkat and B. Boashash, "A High-Resolution Quadratic Time-Frequency Distribution for Multicomponent Signals Analysis," IEEE Transactions on Signal Processing, 49, 2232-2239, 2001.
[42] N. Aimie-Salleh, "Autonomic Function Assessment Tool Using Time-Frequency Analysis of Heart Rate Variability," Master Thesis, Universiti Teknologi Malaysia, Johor Bahru, 2013.
[43] M. B. Malarvili and M. Mesbah, "Newborn Seizure Detection Based on Heart Rate Variability," IEEE Transactions Biomedical Engineering, 56, 2594-2603, 2009.
[44] D. Giri, et al., "Automated diagnosis of Coronary Artery Disease affected patients using LDA, PCA, ICA and Discrete Wavelet Transform " Knowledge-Based Systems, 37, 274-282, 2013.
[45] G. Kheder, et al., "Heart Rate Variability Analysis Using Threshold of Wavelet Package Coefficients " International Journal on Computer Science and Engineering, 1, 131-136, 2009.
[46] R. Acharya, U., et al., "Computer-aided Diagnosis of Diabetic Subjects by Heart Rate Variability Signals Using Discrete Wavelet Transform Method," Knowledge-Based Systems, 81, 56-64, 2015.
[47] B. Oluleye, A. Leisa, J. Leng, and D. Dean, A Genetic Algorithm-Based Feature Selection. International Journal of Electronics Communication and Computer Engineering, 5(4), 899-905, 2014a.
[48] B. Oluleye, A. Leisa, J. Leng, and D. Dean, Zernike Moments and Genetic Algorithm : Tutorial and Application. British Journal of Mathematics & Computer Science 4(15), 2217-2236, 2014b.
[49] N. Akhter, S. Dabhade, N. Bansod, and K. Kale, Feature Selection for Heart Rate Variability Based Biometric Recognition using Genetic Algorithm. In Berretti, S., Thampi, S. M. & Srivastava, P. R. (Ed.) Intelligent Systems Technologies and Applications, Maharashtra, India: Springer International Publishing, 91-101, 2016.
[50] M. S. Mohamad, S. Deris, S. M. Yatim, and M. R. Othman, Feature Selection Method using Genetic Algorithm for the Classification of Small and High Dimension Data. First International Symposium on Information and Communications Technologies. 7-8 October. Putrajaya, Malaysia, 1-4, 2004.
[51] H. Shahamat, and A. A. Pouyan, Feature Selection using Genetic Algorithm for Classification of Schizophrenia using FMRI Data. Journal of AI and Data Mining, 3(1), 30-37, 2015.
[52] S. Ollander, "Wearable Sensor Data Fusion for Human Stress Estimation," Master Master Thesis, Department of Electrical Engineering, Linköping University, Linköping, 2015.
[53] J. Yang, et al., "Feature fusion: parallel strategy vs serial strategy," 36, 1369-1381, 2003.
[54] N. Sharma and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: A survey," Computer Methods and Programs in Biomedicine, 108, 1287-1302, 2012.
[55] J. Zhai and A. Barreto, "Stress Detection in Computer Users Based on Digital Signal Processing of Noninvasive Physiological Variables," in Annual International Conference of the IEEE Engineering in Medicine and Biology Society, New York, 2006, 1355-1358.

[56] Z. B. Siao, et al., "Human emotional stress analysis through time domain electromyogram features," in 2013 IEEE Symposium on Industrial Electronics & Applications, 2013, 172-177.
[57] J.-S. Wang, et al., "A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition," Neurocomputing, 116, 136-143, 2013.
[58] A. Birkhofer, et al., "The deceleration capacity – a new measure of heart rate variability evaluated in patients with schizophrenia and antipsychotic treatment," European Psychiatry, 28,81-86, 2013.
[59] A. Deepak, et al., "Time Domain Measures of Heart Rate Variability during Acute Mental Stress in Type 2 Diabetics: A Case Control Study," National Journal of Physiology, Pharmacy & Pharmacology, 4, 34 – 38, 2014.
[60] A. Muaremi, et al., "Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep," BioNanoScience, 3, 172-183, 2013.
[61] R. Bailon, et al., "Time-Frequency Analysis of Heart Rate Variability During Stress Testing Using "A Priori" Information of Respiratory Frequency," in Computers in Cardiology, Valencia, Spain, 169-172, 2006.
[62] H. M. Seong, et al., "The Analysis of Mental Stress using Time-Frequency Distribution of Heart Rate Variability Signal," in Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Francisco, California, 2004, 283-295.
[63] W. S. Liew, et al., "Classifying Stress From Heart Rate Variability Using Salivary Biomarkers as Reference," IEEE Transactions on Neural Networks and Learning Systems, 27, 2035-2046, 2016.
[64] K. A. Kendall-Tackett, "The health effects of childhood abuse: four pathways by which abuse can influence health," Child Abuse & Neglect 26, 715-29, Jun 2002.
[65] J. P. Shonkoff, et al., "Neuroscience, molecular biology, and the childhood roots of health disparities: Building a new framework for health promotion and disease prevention," JAMA, 301, 2252-2259, 2009.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 325-333 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Performance of a Thermoelectric Powered by Solar Panel for a Large Cooler Box

Mirmanto Mirmanto[1, *], Syahrul Syahrul[1], Made Wirawan[1], I Made Adi Sayoga[1], Agung Tri Wijayanta[2], Ilham Mahyudin[1]

[1]*Jurusan Teknik Mesin, Universitas Mataram, 83125, Indonesia*

[2]*Jurusan Teknik Mesin, Universitas Sebelas Maret, Surakarta, 57126, Indonesia*

ARTICLE INFO

Article history:
Received: 02 November, 2019
Accepted: 19 January, 2020
Online: 07 February, 2020

Keywords:
Thermoelectric
Cooler box
Solar panel
COP

ABSTRACT

An experiment to determine the performance of a thermoelectric powered by solar panel for a large cooler box was carried out. The size of the cooler box tested was 1000 mm x 500 mm x 400 mm and inside the cooler box, a plastic bottle containing 19 liters of water was placed. The thermoelectric hot side was cooled using a mini channel flowed with water. Meanwhile, the cold side of the thermoelectric was connected to the inner heat sink to absorb heat inside the cooler box. The thermoelectric was powered from a battery charged by solar panels directly. There were 2 solar panels, each of which had a power of 100 WP. The cooler box was tested for 6 days in June 2019. The results showed that the lowest temperature on the thermoelectric cold side was around 16.21°C, while the lowest temperature inside the cooler box was 24.25°C. The experimental COPs obtained were ranging from 0.01 to 0.76. Moreover, in general, solar panels were potential as power sources for thermoelectric cooling systems.

1. Introduction

Portable refrigerators are sometimes very necessary, especially in a place far from home or for transportation of goods, blood, and vaccines. For the purposes, thermoelectric cooling machines can be an option because thermoelectric cooling machines are more compact, with no leakage problems, lighter, portable, easier to maintain and more durable as reported in [1- 6]. Nevertheless, the weakness of the thermoelectric is low COP, even less than 1.

There are several ways to raise the COP such as improving the quality of thermoelectric materials, varying the use of thermoelectric techniques, and engineering the dissipation units. The research on thermoelectric materials also has been increased to obtain material that can generate higher COP. But in the term of materials, the thermoelectric development is slow, therefore, attempts to increase the COP through applications have been also tried, one of which is the application to cool beverage cans as reported in [7].

In [7], the authors elucidated that thermoelectric applications to cool beverage cans commonly produced temperatures of 6-8°C within 2 hours. However, the COP of this cooling system was still less than 1, which was 0.856 and could not reach a temperature of 0°C. The authors in [8] investigated the application of thermoelectric modules as a beverage cooler box. They used 2 and 3 thermoelectric modules with a fan or without a fan and a capacity of 34 liters. The beverage cooler box was run for 150 minutes and they obtained a minimum temperature of 14.3°C without water and 16.4°C with 1 liter of water.

In [9], the author examined a thermoelectric refrigerator with 3 thermoelectric modules and power of 51.27 W. The temperature of the refrigerator reached of around 14°C with a drinking water of 1500 ml, while the COP obtained was 0.25. However, even from the engineering side, the application of thermoelectrics is still difficult to obtain high COP. Therefore, this study tries to use a heat dissipation unit that is different from previous studies.

In general, research on the use of thermoelectric cooling modules was carried out using a fin-fan heat sink heat dissipation unit. In [4], authors conducted a study of two different heat dissipation units namely the fin-fan heat sink and the double fan heat pipe. They concluded that with a double fan heat pipe, the cooling box performance including the COP increased. However, the power used by the double fan heat pipe also increased. If the electrical energy consumed by the double fan was taken into account, the heat sink fin-fan unit was superior. Therefore, it is

[*]M. Mirmanto, Jl. Majapahit, no. 62, Mataram, +6282111738971, Email: m.mirmanto@unram.ac.id

necessary to look for other heat dissipation units such as water heat exchangers.

The water heat exchangers such as a water block used for removing heat from the hot side of thermoelectrics had been studied in [10]. Apparently, the water block was still not effective because the ability to transfer heat was still low. Water heat exchangers that might be able to increase COP are mini-channel or mini-pin fin. Why a mini channel is predicted to be able to increase COP? Based on the previous research, mini-channel has been presenting to have a higher heat transfer ability compared to other water heat exchangers as reported in [11]. Similarly, the mini pin fin heat exchanger had been tested to be able to transfer more heat as explained in [12].

Furthermore, COP is also influenced by the size of the refrigerator. A large refrigerator increases the COP and the temperature. Nevertheless, the high refrigerator temperature means that the cooling process is not adequate. Likewise, vice versa, a small refrigerator produces a small COP and a low temperature. However, the effect of refrigerator size is unclear for the refrigerator with a large mass of goods. The more mass in the refrigerator, the higher COP and temperatures as explained in [5]. Table 1 shows several sizes of cooler boxes that have been studied. In Table 1, the largest refrigerator/cooler box size studied is 620 mm x 595 mm x 1565 mm. The refrigerator was examined in [15]. However, they used 8 thermoelectric modules.

Power for thermoelectric refrigerators is generally obtained from conventional electric grids. However, in Indonesia, the price of electricity is still relatively high and the availability of electricity has not been fully distributed to remote areas and small islands yet. Therefore, this study uses electricity sources obtained from solar panels. Even so, this paper does not discuss solar panels in detail because the solar panels are only used as the DC power source.

This research has the advantage of overcoming the urgent problem: (i) the supply of fossil fuels is almost gone, so all electrical equipment including refrigerators may be operated with renewable energy, (ii) introducing new cooling methods that do not require large and expensive equipment, (iii) thermoelectric cooler boxes with mini-channel heat dissipation units have not been widely studied. The purpose of this study is to examine the performance of a large refrigerator using a single thermoelectric module powered using solar panels. The thermoelectric hot side is cooled utilizing a water mini-channel. The characteristics of the refrigerator are represented by the lowest temperatures that can be achieved and the COP. The contribution of this research to the thermoelectric study field is to increase the information concerning thermoelectric usages for cooler boxes.

2. Research Method

2.1. Experimental Set-up

This work was experimental research using a test rig as shown in Figure 1 consisting of a cooler, a thermoelectric, solar panels, water, a mini-channel, a pump, a small radiator, a small tank, a flow meter, a battery, a solar charger, and a data logger National Instrument (NI9714). All temperatures were measured using K type thermocouples calibrated with an accuracy of ± 0.5°C. The electrical power used was measured using a Vichy Vc8145 digital multimeter, while the flow of the water passing through the mini-channel was measured using an FLR1000 flowmeter with an accuracy of ± 0.2 g/s.

When the electric current was supplied to the thermoelectric, the thermoelectric hot side temperature increased drastically. This increased temperature should be maintained so that it did not exceed 40°C, i.e. using water flow. Finally, the heat from the thermoelectric was carried away by the flow of water towards the small radiator. Inside the small radiator, the heat was discharged into the environment through the radiator walls. Then the water returned to the small tank. The electricity used fluctuated because the source was solar panels. When the sun was bright, the power was great and vice versa. However, these fluctuations could be prevented using a battery.

Table 1: Sizes of cooler boxes that had ever been studied, heat dissipation units, and COPs

Reference	Size	Heat dissipation unit	COP
[1]	180 mm x 230 mm x 320 mm	Heat sink with a fan	0.16
[3]	-	Heat sink with a fan	Decrease with difference temperature
[4]	285 mm x 245 mm x 200 mm	Heat sink with a fan, and double fan heat pipe	0.002-0.02
[7]	16 mm x 8 mm x 9 mm	Heat sink with a fan	0.856; 0; 731
[8]	500 mm x 315 mm x 291 mm	Heat sink with a fan	-
[13]	-	Heat sink with a fan	0.43-0.45
[14]	250 mm x 250 mm x 350 mm	Heat sink with a fan	0.65
[15]	620 mm x 595 mm x 1565 mm	Heat sink, big aluminum	0.58
[16]	1. 65 mm x 70 mm x120 mm. 2. 55 mm x 57 mm x 140 mm 3. 100 mm x 100 mm x 230 mm 4. 60 mm x 65 mm 150 mm 5. 80 mm x 75 mm x 170 mm	Heat sink with a fan	-
[17]	-	Heat sink with fan	0.03-0.194

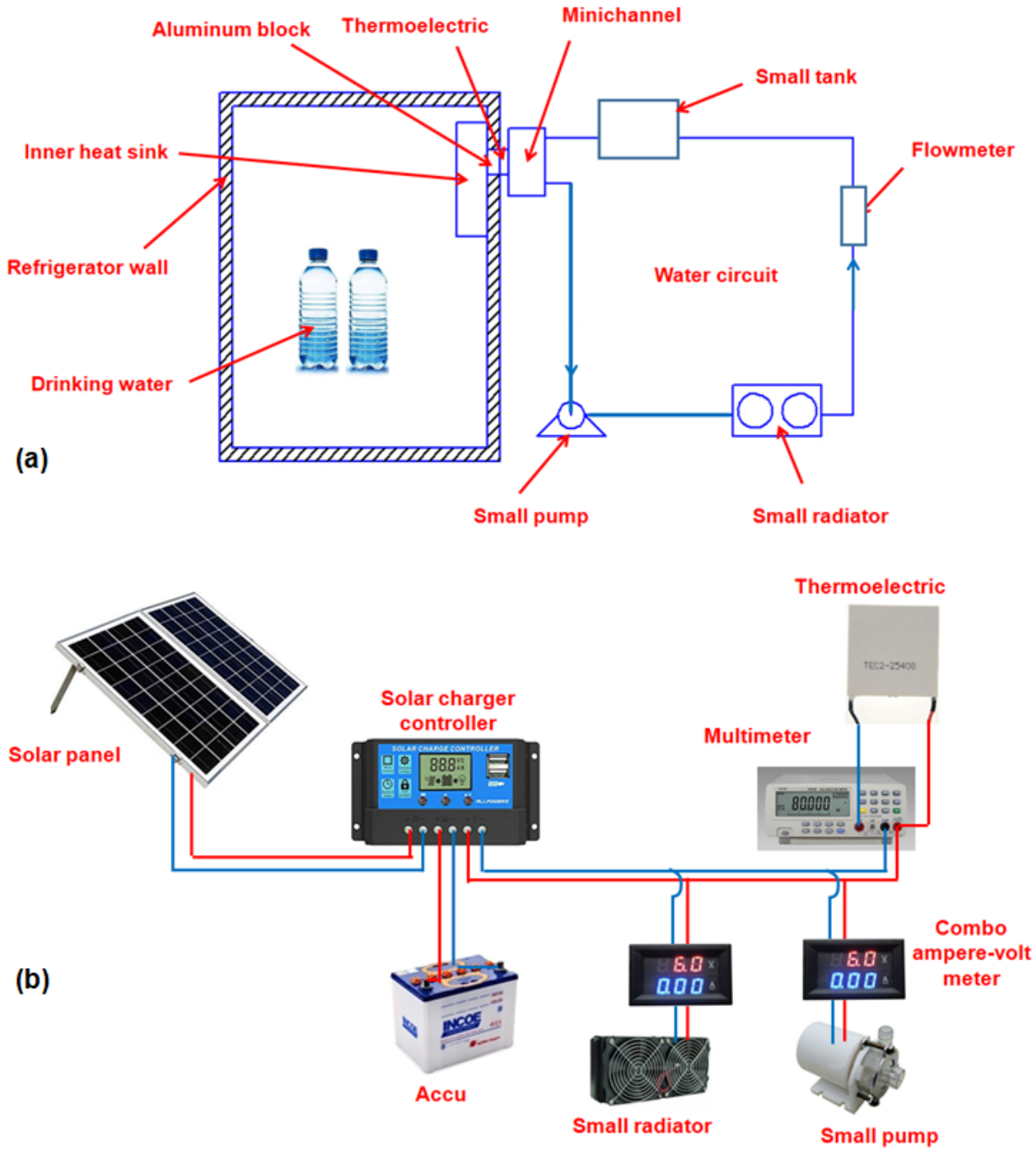


Figure 1: Schematic diagram of the research apparatus; (a) refrigerator and water circuit, (b) electrical circuit.

Thermocouple placements can be seen in Figure 2. The thermocouple locations are denoted using letters a-q. The solar panels used were two pieces with each power of 100WP, and the battery employed was a dry (free maintenance) battery with a voltage of 12 V and a current of 100 AH. However, the analysis here is focused on the performance of the refrigerator only and deep analysis of the mini-channel is not given in this paper.

2.2. Heat Transfer Analysis

The experimental data were analyzed using various equations to determine the performance of the refrigerator. All parameters of the research were analyzed to obtain conclusions. The heat balance in the thermoelectric can be estimated using equation (1) which can be obtained in [14].

$$Q_h = Q_c + P_{in} \tag{1}$$

Q_h is the heat released on the hot side of the thermoelectric (W), Q_c is the heat absorbed on the cold side of the thermoelectric (W) and P_{in} is the input power (W).

In the design of thermoelectric refrigerators, there were two types of heat loads that should be overcome, namely (i) the heat load from the air and the items in the refrigerator, and (ii) the heat

load entering through the walls of the cooler box. The equation to determine the heat load from air and goods can be seen in [4-6, 18-19], which can be expressed in the equation as follows:

$$Q = \frac{\sum_{1}^{n} m_{(i)} c_{p(i)} \left(T_i - T_{i-1}\right)}{t} \quad (2)$$

Q is the heat load and is defined as the total heat coming from the goods inside the cooler box (W). $m_{(i)}$ is the momentary mass (kg), $c_{p(i)}$ is the momentary specific heat of the goods (J/kg K) and $T_{(i)}$ represents the momentary temperature (°C), while t is the total time of the running machine (s). Whereas heat entering through the refrigerator wall that is called conduction can be estimated by the equation that can be found in [20].

$$Q_k = -kA \frac{\sum_{1}^{n} dT_{(i)} t_{(i)}}{dxt} = -\frac{\Delta T_{(i)} t_{(i)}}{Rt} \quad (3)$$

Q_k is the conduction heat transfer rate (W), k is thermal conductivity (W/m°C), A denotes heat transfer area (m²), and $dT_{(i)}/dx$ is the instantaneous temperature gradient over a distance (°C/m). $\Delta T_{(i)}$ is the instantaneous temperature difference of the hot and cold surface temperatures (°C), while R is a thermal resistance (°C/W). Total cooling capacity can be determined by the equation that can be obtained in [18].

$$Q_c = Q_k + Q \quad (4)$$

Q_c is the total cooling load (W), and the COP of the refrigerator can be computed using equation (5) that can be attained in [21]:

$$COP = \frac{Q_c}{P_{in}} = \frac{Q_c}{VI} \quad (5)$$

The P_{in} is all power given to the refrigerator system including the power given to the thermoelectric, to the pump, and to the small radiator. V is the voltage, and I is the current. The thermoelectric hot side was cooled using a mini-channel, in which water flowed. To flow the water, a mini pump was installed on the system, and the pump required electrical power as well. In addition to the pump, the small radiator also required power, because this small radiator utilized two fans, which were driven by DC electricity. So the total power used was the sum of the thermoelectric power, pump power, and small radiator power. *COP* was the comparison of Q_C to total power as shown in equation (5). All power was taken from a battery, and the battery was filled by solar panels.

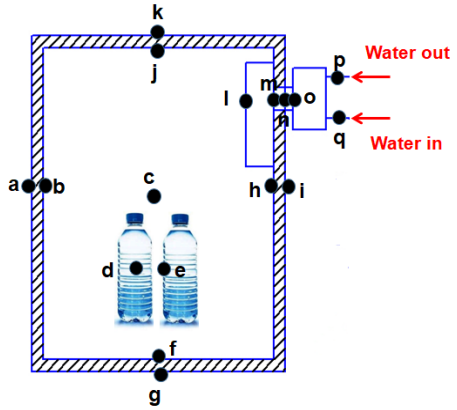


Figure 2: Thermocouple locations are marked with the letter a-q

3. Results and Discussion

The results of the study were presented in the form of graphs and tables. Before the results of the experiments were presented, the recorded temperature results were given in Figure 3 for 6 days. The purpose of presenting this figure was to see the temperature trend and the temperature differences for 6 days.

The thermoelectric hot side temperature increased drastically and then flattened out. This was because heat, Q_h, came out from the thermoelectric hot side. This was also observed by several previous researchers, e.g. [1, 4-6, 14, 22]. Meanwhile, the thermoelectric cold side temperature dropped dramatically and then flattened out. The thermoelectric cold side temperature trend also showed a similar trend, but the room temperature was higher than the thermoelectric cold side temperature. This was because the heat transferred from the room to the cold side of the thermoelectric went through a long connector. In the connector, there could be other heat coming from the surrounding, and then the heat absorbed in the cooler box room was lower than the heat absorbed by the cold side of the thermoelectric.

From the air in the refrigerator, the heat was absorbed by the inner heat sink. From the inner heat sink, the heat went through an aluminum block to arrive at the surface of the thermoelectric cold side. However, at the same time, there was the heat that entered through the refrigerator wall. The heat might come to the connector as well. As a result, the cold side of the thermoelectric was unable to absorb much heat from the refrigerator because it had gained additional heat at the connector.

From the cooler box temperature, water temperature, and bottle temperature, then the total heat that should be removed from inside the cooler box could be estimated. The estimation of heat transfer from air, water, and bottles can be done using equation (2). The results of the experiment showed that the heat taken from the air, from water, and from the bottle decreased with time, and then flattened out, see Figures 4-7. This was because the heat inside the cooler box could not be absorbed anymore after several seconds. The thermoelectric capability was maximal in these conditions. However, such things were also obtained by previous researchers such as [4-6, 14, 23]. The parameter used to indicate the performance of the cooler box was COP. COP could be determined using equation (5). The COP trend decreased with time. In the initial seconds, the COP was very high, then the COP dropped continuously over time, see Figure 10.

This COP trend was also found by some previous researchers such as [13-14, 22, 24-26]. In [26], the authors drained water on the thermoelectric hot side using a mini-channel. However, they used conventional electricity sources and many thermoelectric modules, while in this experiment, the electrical power was taken from solar panels and the thermoelectric module used was only one. How to flow water in [26] was also different from this study.

They flowed water from the reservoir to the pump, to the thermoelectric hot side, to the heat exchanger (small radiator), and back to the reservoir. In this study, the water was flowed from the reservoir to the thermoelectric hot side, to the pump, to the small radiator and back to the reservoir. So the difference was the location of the pump installed on the system/ test rig.

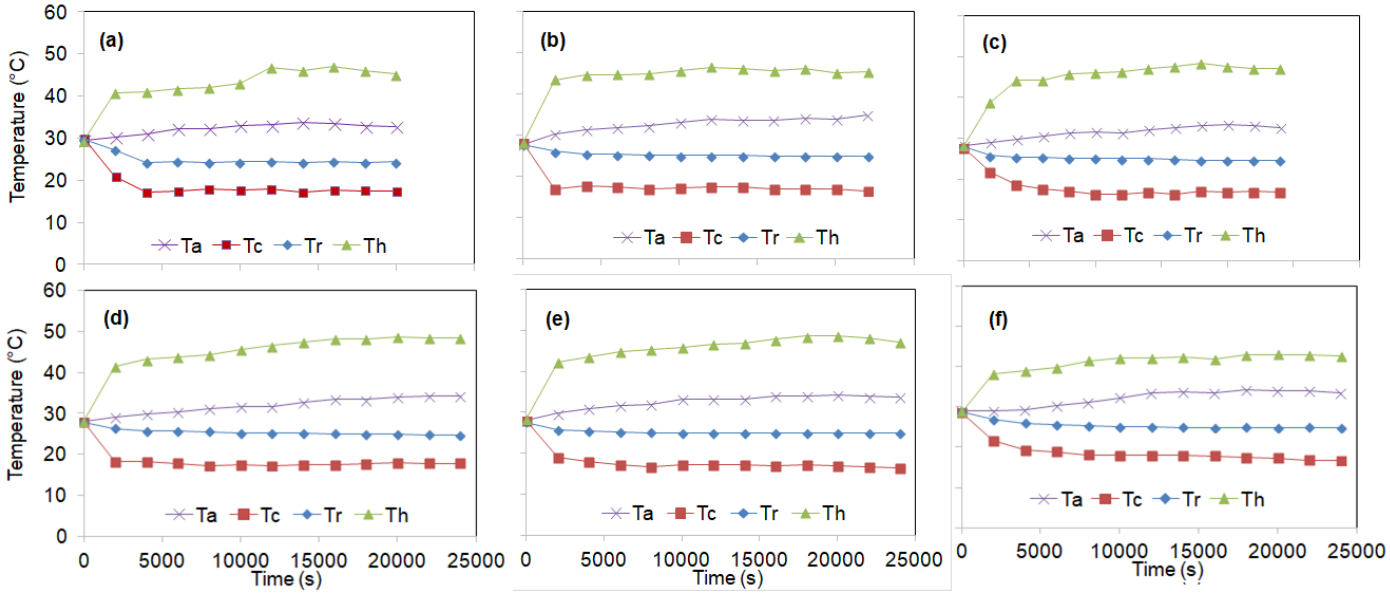


Figure 3: Trend of temperatures with time for 6 days; (a) day 1, (b) day 2, (c) day 3, (d) day 4, (e) day 5, (f) day 6.

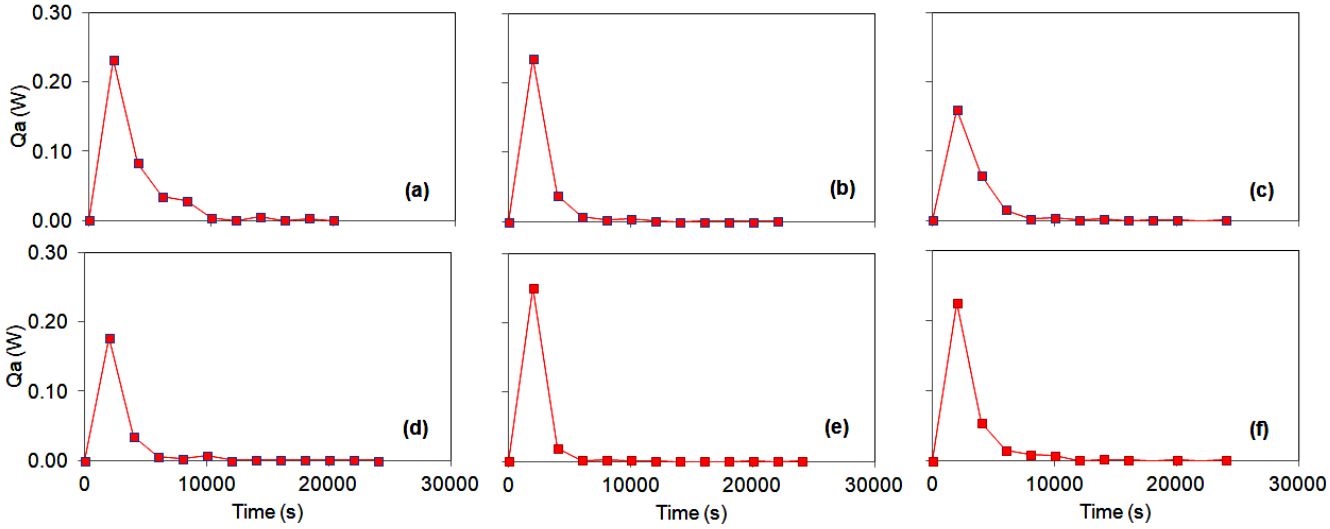


Figure 4: Calculated heat absorbed from the air inside the refrigerator; (a) day 1, (b) day 2, (c) day 3, (d) day 4, (e) day 5, (f) day 6.

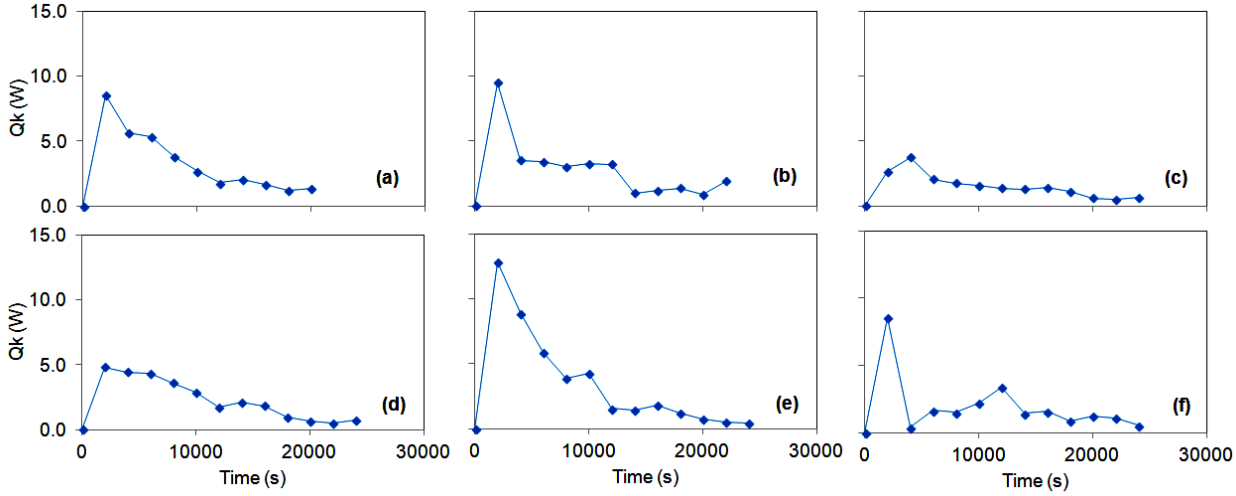


Figure 5: Trends of the conduction heat transfer rate with time; (a) day 1, (b) day 2, (c) day 3, (d) day 4, (e) day 5, and (f) day 6

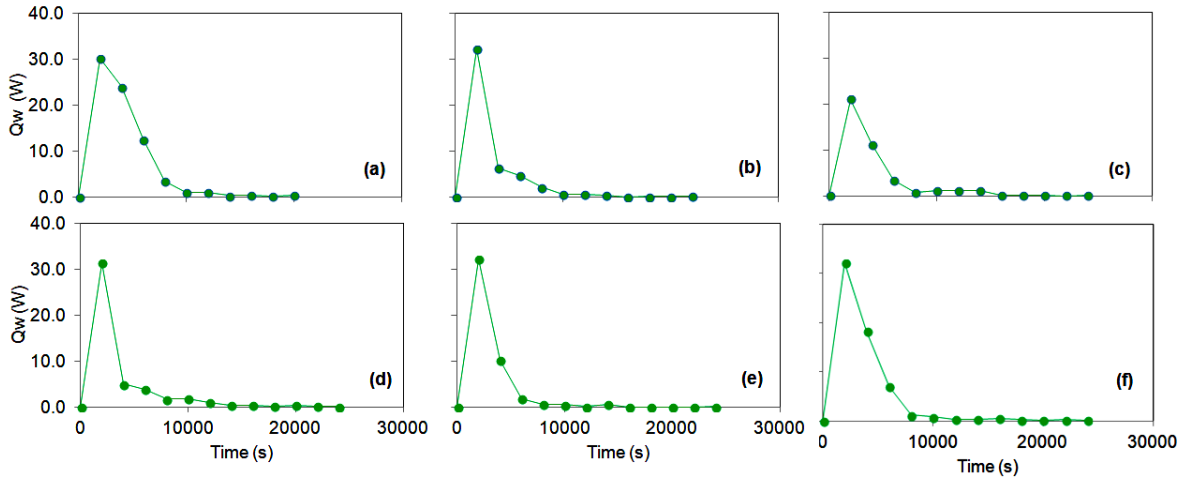


Figure 6: Trends of the heat transfer rate absorbed from the water respect to time; (a) day 1, (b) day 2, (c) day 3, (d) day 4, (e) day 5, and (f) day 6

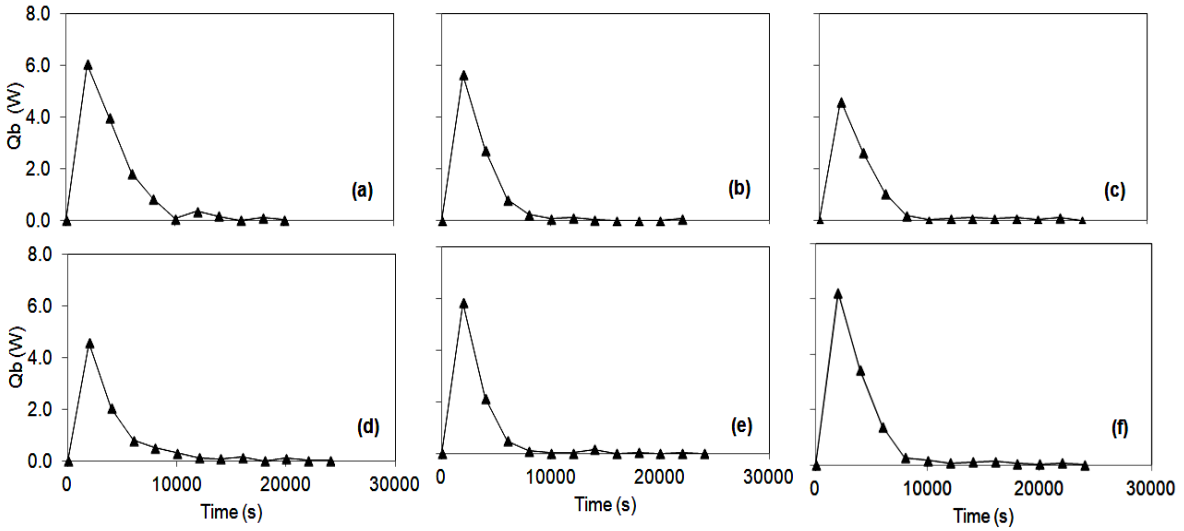


Figure 7: Trends of the heat transfer rate absorbed from the plastic bottle with time; (a) day 1, (b) day 2, (c) day 3, (d) day 4, (e) day 5, and (f) day 6

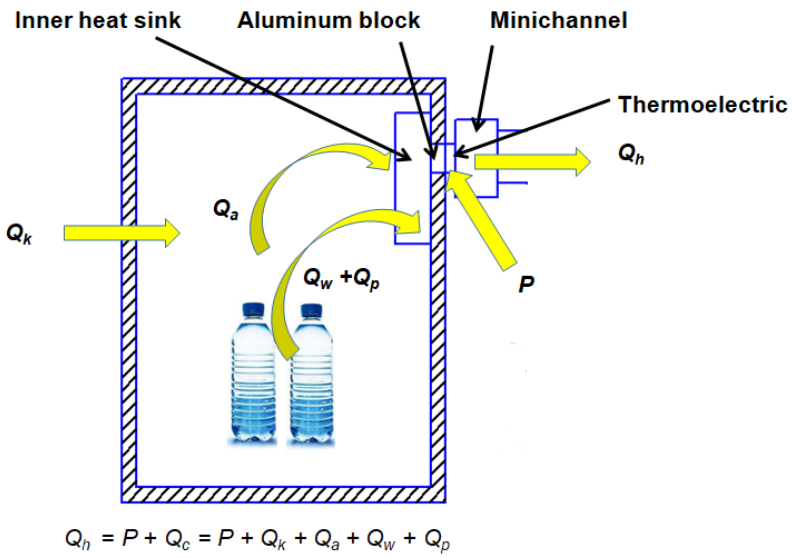


Figure 8: Illustration of heat flows in the refrigerator system for this study

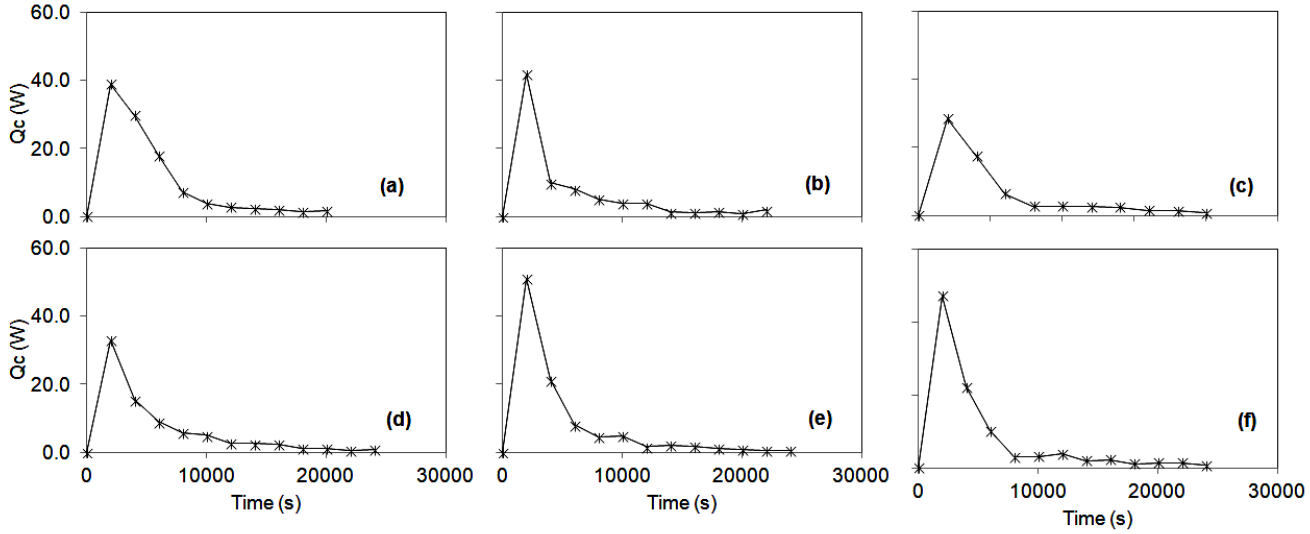


Figure 9: Trends of total heat transfer rate that should be removed from inside the refrigerator with time; (a) day 1, (b) day 2, (c) day 3, (d) day 4, (e) day 5, and (f) day 6

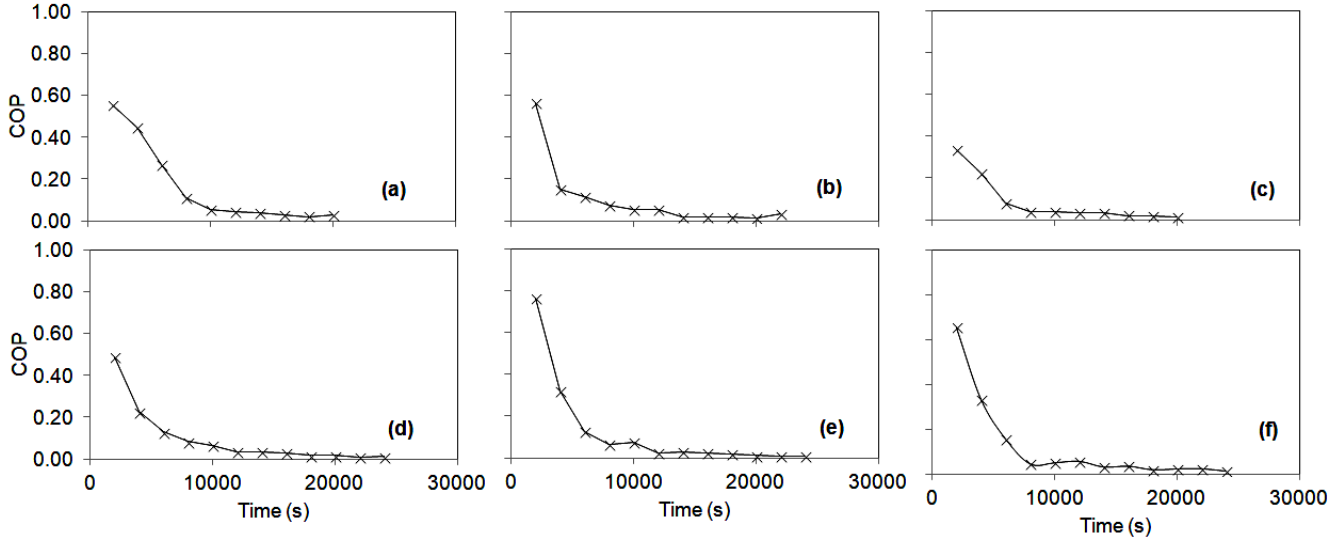


Figure 10: Trends of the COP with time; (a) day 1, (b) day 2, (c) day 3, (d) day 4, (e) day 5, and (f) day 6

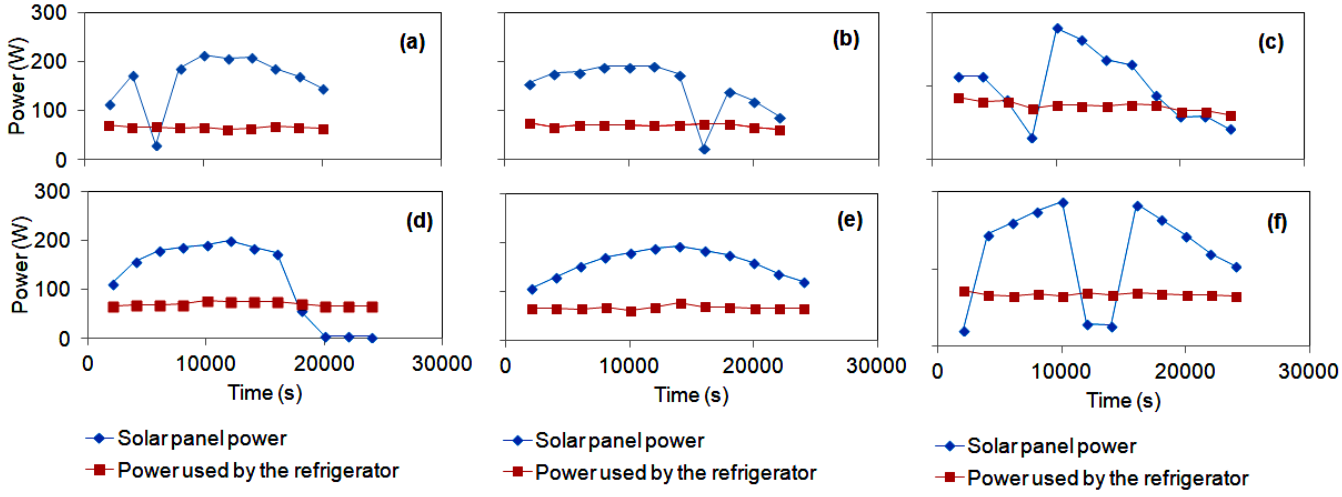


Figure 11: Solar panel power and power used by the refrigerator with time

From the character of the refrigerator mentioned above, the most interesting thing was the relation of the character to the output power of the solar panel. The output power of the solar panels during the study was recorded and shown in Figure 11. The power of solar power increased with time and finally after reaching its peak it dropped again. However, solar panel power was very dependent on sky conditions. If the sky was suddenly cloudy, the solar panel power went down and vice versa. Therefore, the increase in solar panel power was uncertain. Besides that, in this study, the solar panel power was determined as the multiplication of solar power and solar panel area. Furthermore, the multiplication results were then multiplied by the efficiency of solar panels which in general was around 20%. Apparently, fluctuations in solar panel power did not affect the power used by the refrigerator. This was because the output of the solar panel was connected to the battery so that the fluctuations could be muted by the battery. The number of solar panels used was 2 pieces, and each had a power of 100 WP, so the total power was around 200 WP. The power output was apparently less than 200 W as shown in Figure 11.

4. Conclusion

An experimental study to measure the performances of thermoelectric powered using solar panels for cooling a big cooler box was conducted. The cooler box had one thermoelectric module. The air temperature inside the cooler box was still high. One thermoelectric was not adequate for cooling purposes for a big cooler box. The number of thermoelectrics installed should be more than 1 in order to fulfill the cooling purposes. Due to the battery, the fluctuation of the solar panel output became collapse and the power absorbed by the thermoelectric was stable for about 25200 s. In general, the solar panel could be used for powering the thermoelectric cooler box, but one module was not enough for a big box.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to acknowledge the DRPM for the funding through the research PDUPT schema.

Nomenclature

A	cross-sectional area (m²)
c_p	specific heat (J/kg°C)
COP	coefficient of performance
dT	temperature difference (°C)
dx	distance (m)
I	current (A)
m	mass (kg)
Q	heat transfer rate (W)
R	thermal resistance (°C/W)
t	time (s)
T	temperature (°C)
P	power (W)
V	voltage (V)
ΔT	temperature change, temperature difference (°C)

Subscript

a	air
c	cold
h	hot
in	input
k	conduction
p	plastic bottle
w	water

References

[1] S.A. Abdul-Wahab, A. Elkamel, A.M. Al-Damkhi, I.A. Al-Habsi, H.S. Al-Rubai'ey', A.K. Al-Battashi, A.R. Al-Tamimi, K.H. Al-Mamari, M.U. Chutani, "Design and Experimental Investigation of Portable Solar Thermoelectric Refrigerator" Renewable Energy, 34, 30-34, 2009. https://doi.org/10.1016/j.renene.2008.04.026

[2] M.K. Rawat, H. Chattopadhyay, S. Neogi, "A Review on Developments of Thermoelectric Refrigeration and Air Conditioning Systems: A Novel Potential Green Refrigeration and Air Conditioning Technology" Int. J. Engineering Technology and Advanced Engineering, 3(3), 362-367, 2013.

[3] N.J.M. Reddy, "A Low Power, Eco-Friendly Multipurpose Thermoelectric Refrigerator" Front. Energi, 10(1), 79-87, 2016. https://doi.org/10.1007/s11708-015-0380-8

[4] Mirmanto, I.B. Alit, I.M.A. Sayoga, R. Sutanto, Nurchayati, A. Mulyanto, "Experimental Cooler Box Performance Using Two Different Heat Removal Units: A Heat Sink Fin-Fan, and a Double Fan Heat Pipe" Frontier in Heat and Mass Transfer, 10(34), 1-7, 2018. http://dx.doi.org/10.5098/hmt.10.34

[5] M. Mirmanto, I.W. Joniarta, I.M.A. Sayoga, N. Nurpatria, Y.A. Padang, I.G.N.K. Yudhyadi, "Effect of Water Volume on a Thermoelectric Cooler Box Performance" Frontiers in Heat and Mass Transfer, 11(17), 1-9, 2018. http://dx.doi.org/10.5098/hmt.11.17

[6] M. Mirmanto, S. Syahrul, Y. Wirdan, "Experimental performances of a thermoelectric cooler box with thermoelectric position variations" Engineering Science and Technology, an International Journal, 22, 177–184, 2019. https://doi.org/10.1016/j.jestch.2018.09.006

[7] B. Irwin, A. Ruri, "Kaji Eksperimental Kotak Pendingin Minuman Kaleng dengan Termoelektrik Bersumber Dari Arus Dc Kendaraan dalam Rangkaian Seri dan Paralel", *Seminar Nasional Tahunan Teknik XII*, Bandar Lampung, 2013.

[8] A. Aziz, S. Joko, S. Villager, "Aplikasi Modul Pendingin Termoelektrik Sebagai Meia Pendingin Kotak Minuman" Jurnal Teknik Mesin Universitas Riau, 32-38, 2014.

[9] G. Ramdhan, "Sistem Pendinginan pada Kulkas Termoelektrik dengan Variasi Pembebanan" Skripsi, Jurusan Teknik Mesin, Fakultas Teknik, Universitas Mataram, Mataram, NTB, Indonesia, 2016.

[10] D.K. Putra, "Unjuk Kerja Kotak Pendingin Termoelektrik dengan Variasi Laju Aliran Massa Air Pendingi" Skripsi, Jurusan Teknik Mesin, Universitas Mataram, 2018.

[11] J. Hejcik, M. Jicha, "Single-Phase Heat Transfer in Minichannels" EPJ Web of Conference, 67, 1-4, 2014. http://doi.org/10.1051/epjconf/20146702034

[12] T. Yeom, T. Simon, T. Zhang, M. Zhang, M. North, T. Cui, "Enhanced Heat Transfer of Heat Sink Channels With Micro Pin Fin Roughened Walls" Int. J. of Heat and Mass Transfer, 92, 617–627, 2016. https://doi.org/10.1016/j.ijheatmasstransfer.2015.09.014

[13] M. Gillott, L. Jiang, S. Riffat, "An Investigation of Thermoelectric Cooling Devices for Small-Scale Space Conditioning Applications in Buildings" International Journal Of Energy Research, 34, 776-786, 2010. http://doi.org/10.1002/er.1591

[14] S. Jugsujinda, A. Vora-ud, T. Seetawan, "Analysing of thermoelectric refrigerator performances" Procedia Engineering, 8, 154-159, 2011. https://doi.org/10.1016/j.proeng.2011.03.028

[15] K. Atik, Y. Yildiz, "An Experimental Investigation of a Domestic Type Solar TE Cooler" Energy Sources, Part A, 34, 645–653, 2012. http://dx.doi.org/10.1080/15567031003627989

[16] A. Aziz, R.I. Mainil, A.K. Mainil, Syafri, M.F. Syukrillah, "Design of Portable Beverage Cooler Using One Stage Thermoelectric Cooler (TEC) Module" Aceh Int. J. Sci. Technol., 7(1), 29-36, 2017. http://doi.org/10.13170/aijst.6.1.5427

[17] D. Suryawanshi, N. Pokharkar, V. Pokale, A. Walgude, "Design and Fabrication of Thermoelectric Refrigerator for Liquid Cooling by Automatic

Temperature Micro-Controller" International Journal of Science Technology & Engineering, **3**, 21-27, 2016.

[18] F. Incropera, D.P. DeWitt, T.L. Bergman, A.S. Lavine, "Fundamentals of Heat and Mass Transfer, 6th Edition, John Wiley & Sons, Hoboken, 2006.

[19] H.P. Banjarnahor, "Analisis Laju Pendinginan pada Kulkas Thermoelektrik Super Cooler Dibandingkan Sistem Pendingin Konvensional Menggunakan Gas Freon" Skripsi, USU Institutional Repository, Universitas Sumatera Utara, 2016.

[20] A. Waleed et al., "Solar (PV) Water Irrigation System with Wireless Control," in 2019 International Symposium on Recent Advances in Electrical Engineering (RAEE), 2019, vol. 4, pp. 1–4.

[21] Y.A. Cengel, M.A. Boles, "Thermodynamics an Engineering Approach", 6th Ed., McGraw-Hill Higher Education, Boston, 2008.

[22] S.M. Hafis, M.J.M. Ridzuan, A.Z.A. Firdaus, S.M. Shahril, R.N. Farahana, C.A. Chong, "COP Improvement of Thermoelectric Cooler Through the Optimization of Heat Dissipation System" Applied Mechanics and Materials, 554, 241-245, 2014. https://doi.org/10.4028/www.scientific.net/AMM.554.241

[23] H. Ananta, Mirmanto, Y.A. Padang, "Unjuk Kerja Kulkas Termoelektrik dengan Rangkaian Seri dan Parallel pada Beban Air 1500 ml" Dinamika Teknik Mesin, 7(2), 80-86, 2017. https://doi.org/10.29303/dtm.v7i2.157

[24] B.J. Huang, C.J. Chin, C.L. Duang, "A Design Method of Thermoelectric Cooler" International Journal of Refrigeration, 23, 208-218, 2000. https://doi.org/10.1016/S0140-7007(99)00046-8

[25] H. A. Raza, M. T. Riaz, M. Idrees, M. M. Afzal, W. S. Hashmi, and M. W. Asif, "Analysis the effect of 500kv High-Voltage Power Transmission Line on the Output Efficiency of Solar-Panels," in 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 2019, pp. 1–6.

[26] M. Gökçek, F. Şahin, "Experimental Performance Investigation of Minichannel Water Cooled-Thermoelectric Refrigerator", Case Studies in Thermal Engineering, 10, 54-62, 2017. https://doi.org/10.1016/j.csite.2017.03.004

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 334-339 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Skin Tissue Oxygen Saturation Prediction: A Comparison Study of Artificial Intelligence Techniques

Audrey Huong[1,*], Xavier Ngu[2]

[1]*Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, Malaysia*

[2]*Centre for Applied Electromagnetics, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, Malaysia*

ARTICLE INFO

Article history:
Received: 11 November, 2019
Accepted: 27 December, 2019
Online: 07 February, 2020

Keywords:
Skin tissue oxygen
Partial Least Square
K-Nearest Neighbor
Neural Network

ABSTRACT

Noninvasive measurement of skin tissue oxygen saturation, S_tO_2, is of interest especially in the studies of wound healing and detection of vascular diseases. This work aims to compare Partial Least Square (PLS) regression, K-Nearest Neighbor (KNN) and Artificial Neural Network (ANN) technique in the prediction of S_tO_2 using spectral data obtained from Monte Carlo simulations. We found a good performance of PLS technique with mean and standard deviation (SD) of the absolute prediction errors given by 2(1.37) % for noisy data with signal-to-noise ratio (SNR) of 30 dB. The errors increased to 4.67(3.87) % under SNR of 20 dB. This is followed by KNN technique, which calculated errors varied from 4.28(3.58) % to 7.61(6.08) % with a decrease in SNR. Meanwhile ANN produced large errors in its prediction ranging from 4 to 9 % for the considered SNR values. This work concluded PLS is suitable for use in selecting significant spectral information for the development of prediction models using artificial intelligent (AI) system or a hybrid combination of AIs to achieve a higher accuracy.

1. Introduction

Tissue oxygen saturation is a clinical indicator for assessment and diagnosis of many illnesses such as peripheral vascular diseases [1], organ viability [2], hypoxic-ischemia conditions [3], hemorrhagic shock [4], in studies related to cancers [5] and muscle oxidative metabolism [6], and for management of wounds (burns) and ulcers [5, 7]. In this context, sufficient tissue oxygen pressure combined with adequate supply of oxygen and nutrients through functioning microcirculatory system are more likely to guarantee cell survival.

The current state of the arts of skin tissue oxygen consumption measurement includes transcutaneous oxygen tension ($TcPO_2$). Despite its noninvasive and high accuracy attributes, this equipment required mild heating at the local site in its operation to promote blood perfusion prior to the measurement. Some of the other drawbacks include long measurement time, and sensitivity of its performance to factors such as the person's stress level, body temperature, medication and local inflammation [8]. Systems that take advantage of physical and optical properties of blood for noninvasive and noncontact measurement of blood oxygen saturation are of increasing demand; the said systems included spectroscopy techniques (optical, time and frequency domains) and clinical imaging methods such as Magnetic Resonance Imaging [9,10], and Photoacoustic techniques [11]. Among the spectroscopy approaches, hyperspectral imaging technique has emerged as a popular choice owing to its inherent ability to produce high resolution data in both spectral and spatial dimensions. Even though this system collects images at a shallow depth of field, it involves less complex technologies. With regard to its application in tissue oxygen measurement using light of visible range, it is able to spectrally resolve information carrying light backscattered from cutaneous tissue. The latter can be used in prediction of haemoglobin concentration and tissue oxygen saturation using either an inverse model, or computer aided, or computational intelligence system [12, 13]. Since the system interrogates blood in dermal capillaries, arterioles and venules [14], the predicted value is often taken as the average of oxygen saturation within the cutaneous microenvironment, S_tO_2. Previous works [15, 16] from this laboratory found that the range of predictions is dependent on the system configuration and architecture, and the wavelengths used. Meanwhile the use of Artificial Intelligence (AI) techniques has received a great deal of attention in this arena of study over the past years due to its ability to learn by itself through a large set of data, events and statistics before making decision. This capacity makes it a highly sought-

*Audrey Huong, Email: audrey@uthm.edu.my

after means for classification, prediction and forecasting tasks in many fields (for a review, see [17, 18]). In addition, much less efforts and time is needed to spend on improving the formulations of inverse models and look-up table.

The use of AI in medicine is well recognized to provide better quality medical health care services by reinforcing diagnosis with prediction and identifying definitive intervention, thus reducing morbidity and mortality case, and healthcare cost. Among the applications of AI include diagnosis and prognosis of cancers, illness and disease, to aid in decision makings and for physiotherapy practice [19, 20]. The examples of AI used for clinical screening and imaging tasks are such as optimization algorithms (see ref. [21] for a comprehensive review), neural network [22], and computer aided diagnosis [23].

To the best of the authors' knowledge the use of AI, and the study of the performance of different AI techniques, in the predictions of S_tO_2 using an optical system is few and far between. The further advancement in the integration of these technologies is impeded by insufficiency of data available for the training of the AI system, prompting the work by [24] that used Monte Carlo (MC) platform for training of artificial neural networks for future optical diagnostics and sensing. This work is following in the footsteps of [24], and our aim is to compare the performance of different AI techniques for the prediction of S_tO_2 using MC simulation data.

2. Material and Methods

2.1. Monte Carlo (MC) simulations of light propagation

The two layered human skin model, the illumination and detection systems used for simulations of light propagated in skin shown in Figure 1 were previously presented in [22]. These simulation data are used here again owing to the known medium S_tO_2 for evaluations of performance of prediction techniques described in the following. The simulation data for medium percent S_tO_2 ranging from 0 % to 100 % at a step resolution of 5 % were added with Gaussian White Noise (AWGN function in MATLAB 2018a) to produce signals with signal-to-noise ratio (SNR) of 20 and 30 dB. For each SNR and S_tO_2 level, 10 sets of random noise were incorporated into the simulations to give a total of 210 data for each SNR level, among which 70 % (n = 147) were randomly chosen for the training of the model, while the remaining (n = 63) were used as the testing data.

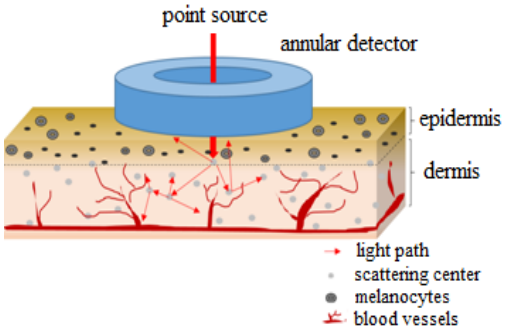


Figure 1: MC simulations of light propagation in a two-layered skin medium

2.2. Features selection

Since light absorption in skin varied with oxygen binding status of hemoglobins (i.e. oxyhemoglobin and deoxyhemoglobin concentration), by examining the absorption differences in the spectra of these hemoglobin components one can deduce the required S_tO_2 parameter. The simulated spectral data in section 2.1 were used as the input data to the artificial intelligence techniques. This work considered signals across the wavelengths ranged from 520 nm to 650 nm where changes in hemoglobin absorptivity signatures with S_tO_2 are significant (image can be found in [25]).

2.3. Tissue oxygen saturation prediction tools

In the present study, we investigated three different machine learning techniques: artificial neural network (ANN), partial least square (PLS) and K-nearest neighbor (KNN) in the estimation of the required percent S_tO_2 value. The error, ε, (in unit of %) in the predicted value, $S_tO_{2(p)}$, is defined as:

$$\varepsilon = S_tO_{2(p)} - S_tO_2' \quad (1)$$

where S_tO_2' is the ground truth value. The prediction tools presented in the following used MATLAB (version 2018b) software in their implementation due to its comprehensive multivariate analysis toolboxes.

2.3.1. Artificial Neural Network (ANN)

This work used *nnstart* toolbox to generate an ANN feedforward model from the simulation data. We have chosen 10 hidden neuron layers, and the fitting of the input-output relationship was by using Levenberg-Marquardt (LM) backpropagation model to update weights in minimizing sum of square error functions. This ANN model was modified from the prior work in [22], wherein 'while' loop structure was adopted in obtaining a neural network with acceptable performance during its training. This is with the implementation of arbitrarily chosen MSE threshold value of 24 and 15, respectively, for dataset of SNR = 20 and 30 dB.

2.3.2. Partial Least Square Regression (PLS)

PLS regression analysis was used here to develop a prediction model. This technique predicts the response, Y, for a set of data by extracting latent variables from sampled observed variables, X, as followed:

$$\begin{aligned} X &= T * P' + E_0 \\ Y &= U * Q' + F_0 \end{aligned} \quad (2)$$

where T and U are latent variables, P' and Q' are loading matrices and E_0 and F_0 are residuals. The latent variables are related via inner relation coefficient, B, and residual, F_1, as followed:

$$U = TB + F_1. \quad (3)$$

This study used *plsregress* function to project high dimensionalities feature vectors (X and Y) onto a subspace to give an overall partial regression model given by:

$$\tilde{Y} = X * (P * B * Q') + F. \quad (4)$$

The $P * B * Q'$ vector in (4) that simplified the regression problem was used to predict the most probable percent S_tO_2 value.

We have chosen latent factor of 3 for demonstration purpose. It must also be mentioned that the present work found no improvement in the prediction accuracy using other factor number.

2.3.3. K-Nearest Neighbor (KNN)

The KNN classification model for predictors, X, and response, Y, was generated using *fitcknn*. This study considered exhaustive search algorithm to find the nearest neighbors by calculating the distances from all points in X to each point in Y. The distance metric that was considered in the present study is Minkowski distance. The K value was chosen as 11 based on the size of the datasets following the discussion in [26]. The trained classifier in (5) was then used to predict the class of the testing data.

$$mdl = fitcknn\,(X, Y). \tag{5}$$

3. Results and analysis

The changes in light attenuation spectrum following variations in the reference S_tO_2 (i.e. ground truth values used in the simulations) are shown in Figure 2. The plots showed example changes in the noise corrupted signals of the considered SNR. These spectral data across wavelength range of 520 – 650 nm and a resolution of 1 nm were used for both training and evaluation of the considered models using a constant random seed for comparable results. The data were separated into training and testing sets with a percentage of 70:30. Figure 3 showed the three-dimensional plots of attenuations from training set for the reference of our readers. These plots also revealed changes in the value range and distribution with noise level. The mean (SD) of the absolute ε calculated from (1) for percent value predicted by ANN, PLS and KNN technique for spectrum of different reference S_tO_2 shown in Table 1 is calculated as 8.76 (6.82) %, 4.67 (3.88) % and 7.62 (6.08) %, and 4.77 (3.96) %, 2 (1.37) % and 4.28(3.57) % for dataset of SNR of 20 dB and 30 dB, respectively. The computation time (averaged from three consecutive runs) spent in the training of the machine learning techniques described in section 2.3.1-2.3.3 performed on an ASUS laptop with 64-bit window 7, Intel® Core™ Duo T6500 CPU @2.10 GHz are given by 1.38 and 1.8 seconds for PLS and KNN, respectively. A longer average time was needed for ANN with the time taken recorded as 81.5 and 39.1 seconds in the training of data of SNR = 20 and 30 dB, respectively.

4. Discussion

Figure 2 showed substantial decrease in the noise level on the spectra for signals of SNR 30 dB as compared to that of 20 dB. This decrease in noise level yields an overall improvement in the prediction accuracy in Table 1 with an average decrease in ε value calculated as 3.33(**0.66**) %. The lowest mean prediction error was observed for values given by PLS technique; this is followed by KNN and ANN. While both ANN and PLS produced regressed value in their predictions, categorical values according to the range of target values used during the training process were produced by the KNN. The main reason PLS technique outperformed its counterparts in the validation phase is likely due to the accentuation of the variations in the absorptivity through the projection of data to a lower dimensional space. An investigation into the correlation coefficients obtained from the regression of centre variables X and Y in (2) using three latent factors revealed a strong positive correlation at wavelength 580 nm, followed by that of 538 nm under SNR = 30 dB. These peaks increasingly diminished under SNR 20 dB.

Since similar distribution of attenuation values with changes in S_tO_2 was observed in Figure 2, and the features used in the training of the considered AI models discussed in section 2.2 are of full spectral information (attenuations across wavelength range 520 –

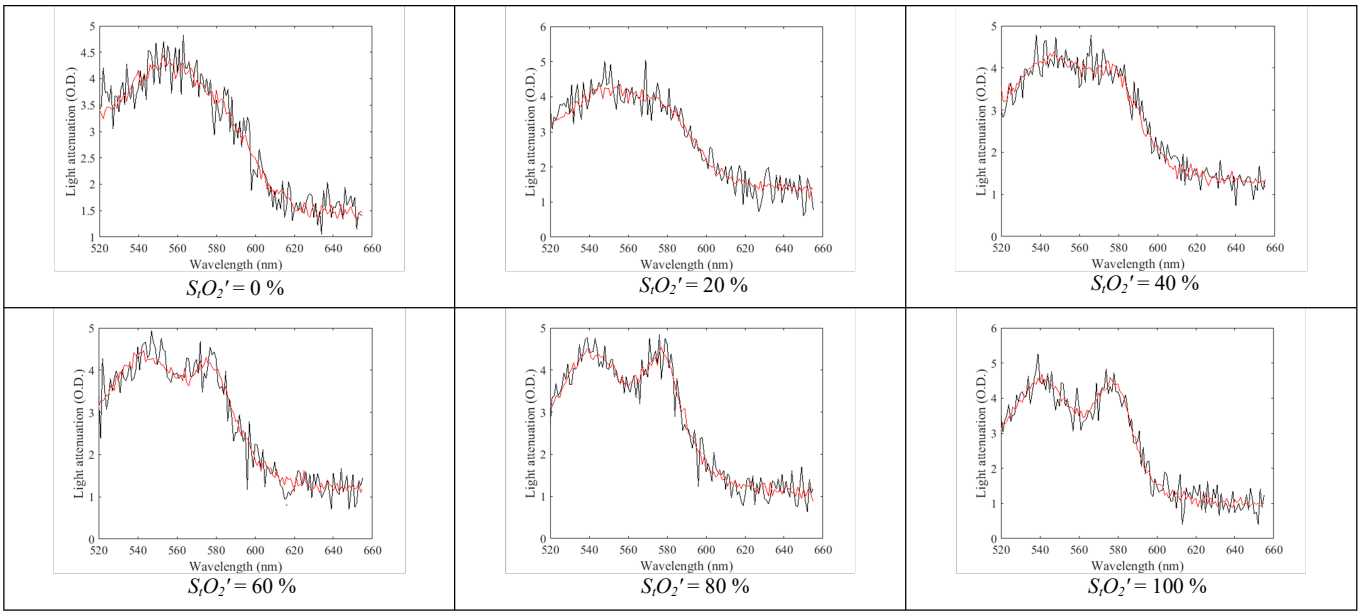
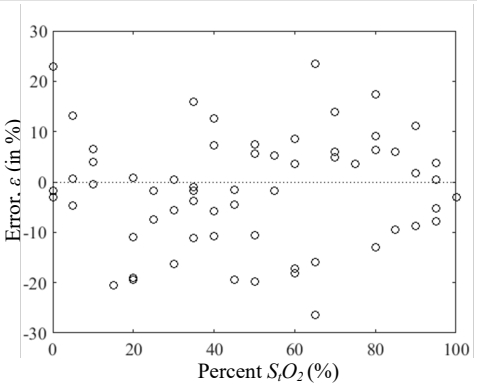
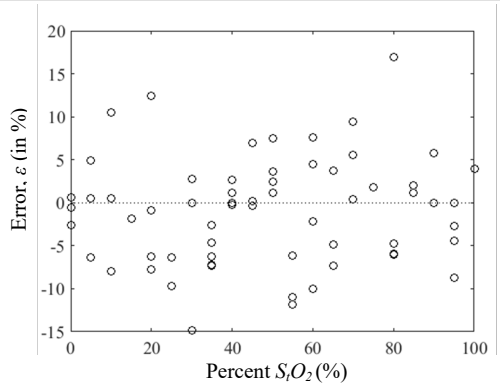
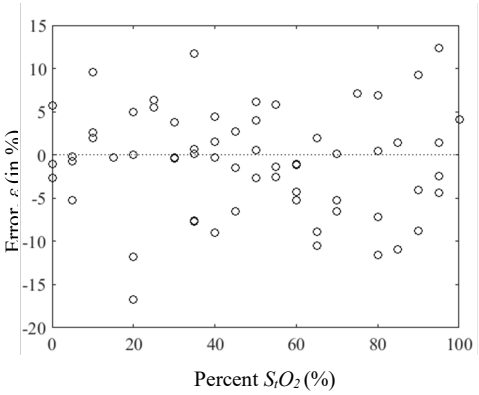
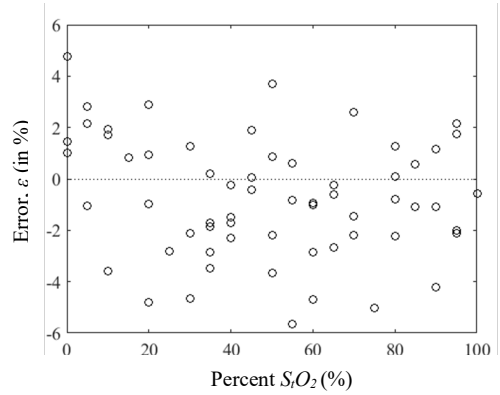
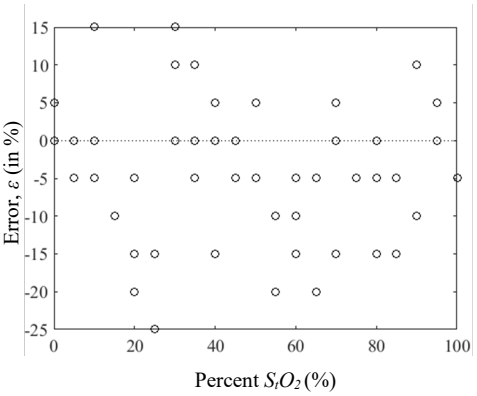
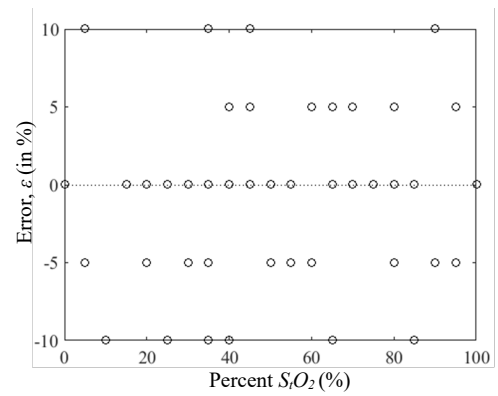


Figure 2: Noise corrupted wavelength dependent light attenuation with Signal to Noise (SNR) of 20 dB (dark lines) and 30 dB (red lines) under different S_tO_2 condition.

Table 1: Error, ε, in the value predicted for testing dataset of SNR 20 dB (Left) and 30 dB (Right) using (from top to bottom) Artificial neural network (ANN), Partial Least Square (PLS) and K-Nearest Neighbor (KNN).

ANN		
PLS		
KNN		

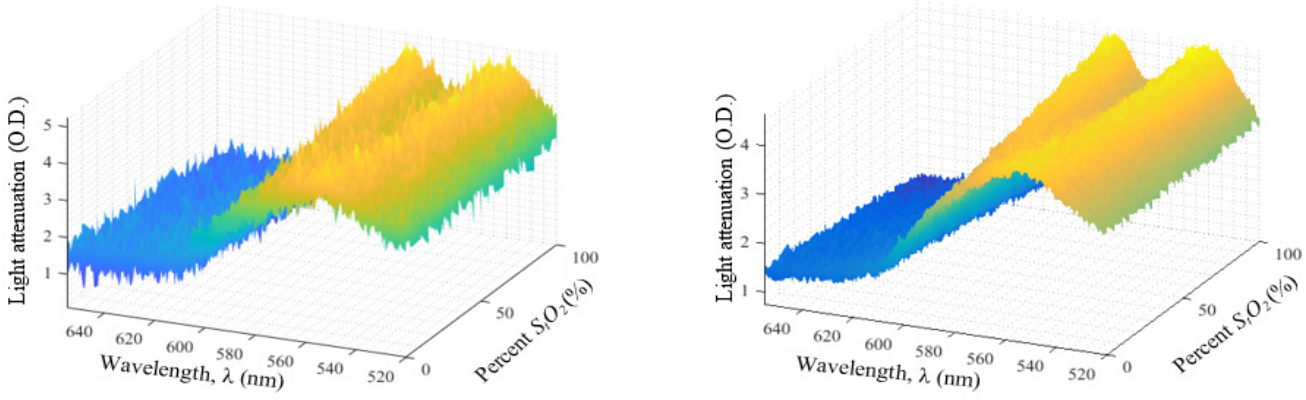


Figure 3. Three-dimensional surface plots of light attenuation values for SNR of (left) 20 and (right) 30 dB.

650 nm at a step resolution of 1 nm), this produces a poor network and association on the relationship between input and target values. The prediction errors can be minimized by considering features that are of utmost significance, such as those within the presence of troughs and crest across wavelengths 540 to 570 nm. An example of previous works from this laboratory in [22] extracted and considered the important features (i.e. magnitude and wavelength of the hemoglobin peaks) in the prediction work, hence minimizing errors from overfitting.

Meanwhile the KNN technique is primarily based on the location (i.e. magnitude) of these attenuation values, large difference in the values with changes in S_tO_2 would be desirable to yield better accuracy result. For this reason, relatively consistent errors were observed in Table 1. We hypothesized that this technique could be successfully applied to the classification of skin pigmentations.

It must be mentioned that the use of two-layered homogeneous medium is insufficient to represent human skin tissues since there are differences in optical properties of heterogeneous sub-layers in skin as widely reported in earlier works [27, 28]. Nonetheless the findings from this study would be useful when measurement data are readily available.

5. Conclusion

This work investigated the performance of different AI techniques in the prediction of medium S_tO_2 using MC technique. The significant findings of this study include the validation of the importance of hemoglobin signatures in the prediction of the required S_tO_2 value. The notably better accuracy in the values predicted using PLS regression technique is owing to its ability to distinguish the hemoglobin peaks using latent factor of 3. Meanwhile the shortcoming of KNN technique in the prediction work is largely contributed by the high similarity in the range and distribution of light attenuation values making the classification process complex and difficult. In addition, the large spectral information used in ANN technique resulted in over-fitting problem and produced comparatively larger errors. This work concluded that PLS regression performed well in the prediction of S_tO_2, and this technique may also be suitably used to select the distinctive signatures of spectral prior to the development of a prediction model using other AI technique.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to acknowledge Research Management Center (RMC) Universiti Tun Hussein Onn Malaysia for funding this study under research fund Vot E15501.

References

[1] R. M. Siao, M. J. So, M. H. Gomez, "Pulse oximetry as a screening test for hemodynamically significant lower extremity pheripheral artery disease in adults with type 2 diabetes mellitus" Journal of the ASEAN Federation of Endocrine Societies, **33**(2), 1-20, 2018. https://doi.org/10.15605/jafes.033.02.04

[2] N. T. Clancy, S. Saso, D. Stoyanov, V. Sauvage, D. J. Corless, M. Boyd, D. E. Noakes, M. Y. Thum, S. G. Maghami, J. R. Smith, D. S. Elson, "Multispectral imaging of organ viability during uterine transplantation surgery in rabbits and sheep" Journal of Biomedical Optics, 21(10), 1060061-7, 2016. https://doi.org/10.1117/1.JBO.21.10.106006

[3] S. Akter, S. Kawauchi, S. Sato, S. Aosasa, J. Yamamoto, I. Nishidate, "In vivo imaging of hepatic hemodynamics and light scattering property during ischemia-reperfusion in rats based on spectrocolorimetry" Biomedical Optics Express, 8(2), 972-992, 2017. https://doi.org/10.1364/BOE.8.000974

[4] C. Carlile, C. E. Wade, M. S. Baraniuk, J. B. Holcomb, L. J. Moore, "Evaluation of S_tO_2 tissue perfusion monitoring as a tool to predict the need for lifesaving interventions in trauma patients" The American Journal of Surgery, 210(6), 1070-5, 2015. https://10.1016/j.amjsurg.2015.08.006

[5] R. Rutkowski, M. Schuster, J. Unger, C. Seebauer, H. R. Metelmann, Th.v. Woedtke, K. D. Weltmann, G. Daeschlein, "Hyperspectral imaging for in vivo monitoring of cold atmospheric plasma effects on microcirculation in treatment of head and neck cancer and wound healing" Clinical Plasma Medicine, 7(8) 52-57, 2017. https://doi.org/10.1016/j.cpme.2017.09.002

[6] R. Re, I. Pirovano, D. Contini, L. Spinelli, A. Torricelli, "Time domain near infrared spectroscopy device for monitoring muscle oxidative metabolism: Custom probe and in vivo applications" Sensors, 18(1), 264, 2018. https://doi.org/10.3390/s18010264

[7] S. Zhang, S. Gnyawali, J. Huang, W. Ren, G. Gordillo, C. K. Sen, R. Xu, "Multimodal imaging of cutaneous wound tissue" Journal of Biomedical Optics, 20(1), 016016, 2015. https://doi.org/10.1117/1.JBO.20.1.016016

[8] A. Y. Mousa, J. L. Ballard, Transcutaneous Oxygen Tension: Principles and Applications, Springer Link, 2017

[9] M. B. Vestergaard, U. Lindberg, N. J. Aachmann-Andersen, K. Lisbjerg, S. J. Christensen, I. Law, P. Rasmussen, N. V. Olsen, H. B. W. Larsson, "Acute hypoxia increases the cerebral metabolic rate- a magnetic resonance imaging study" Journal of Cerebral Blood Flow & Metabolism, 36(6), 1046-1058, 2016. https:// 10.1177/0271678X15606460

[10] Z. Deng, C. Li, "Noninvasively measuring oxygen saturation of human fnger-joint vessels by multi-transducer functional photoacoustic tomography" Journal of Biomedical Optics, 21(6), 0610091-5, 2016. doi: 10.1117/1.JBO.21.6.061009

[11] M. Barberio, F. Longo, C. Fiorillo, B. Seeliger, P. Mascagni, V. Agnus, V. Lindner, B. Geny, A.L. Charles, I. Gockel, M. Worreth, A. Saadi, J. Marescaux, M. Diana, "Hyperspectral enhanced reality a physiology-based surgical guidance tool" Surgical Endoscopy, 1-9, 2019. https://doi.org/10.1007/s00464-019-06959-9

[12] Q. Wang, Q. Li, M. Zhou, Z. Sun, H. Liu, Y. Wang, "A hyperspectral vessel image registration method for blood oxygenation mapping" PLOS ONE, 2017. https://doi.org/10.1371/journal.pone.0178499

[13] W. Feng, R. Shi, C. Zhang, T. Yu, D. Zhu, "Lookup table based inverse model for mapping oxygen concentration of cutaneous microvessels using hyperspectral imaging" Optics Express, 25(4), 3481-3495, 2017. doi:10.1364/oe.25.003481

[14] A. Vogel, V. V. Chernomordik, J. D. Riley, M. Hassan, F. Amyot, B. Dasgeb, S. G. Demos, R. Pursley, R. Little, R. Yarchoan, Y. Tao, A. H. Gandjbakhche, "Using noninvasive multispectral imaging to quantitatively assess tissue vasculature" Journal of Biomedical Optics, 12(5), 051604 -,2007. https://doi.org/10.1117/1.2801718

[15] S. P. Philimon, A. K. C. Huong, X. T. I. Ngu, "Investigation of spatial resolution dependent variability in transcutaneous oxygen saturation using point spectroscopy system" in 2017 IOP Conf. Ser.: Mater. Sci. Eng. 226, 012122, 2017.

[16] S. P. Philimon, A. K. C. Huong, X. T. I. Ngu, "An alternative wavelength range for noninvasive assessment of wound tissue oxygenation status" International Journal of Engineering and Technology, 7 (4.26), 73-33, 2018. DOI: 10.14419/ijet.v7i4.26.22141.

[17] R. Liu, B. Yang, E. Zio, X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review" Mechanical Systems and Signal Processing, 108, 33- 47, 2018. https://doi.org/10.1016/j.ymssp.2018.02.016

[18] M. I. Jordan, T. M. Mitchell, "Machine leaning: Trends, perpecrtives, and prospects" Science, 349 (6245), 255 – 260, 2015. DOI: 10.1126/science.aaa8415

[19] H. Akbari, L. V. Halig, H. Zhang, D. Wang, Z. G. Chen, B. Fei, "Detection of cancer metastasis using a novel macroscopic hyperspectral method" in Proceedings of SPIE--the International Society for Optical Engineering, 8317, 831711. doi:10.1117/12.912026

[20] V. A. Convertino, S. L. Moulton, G. Z. Grudic, C. A. Rickards, C. Hinojosa-Laborde, R. T. Gerhardt, L. H. Blackbourne, K. L. Ryan, "Use of advanced machine learning techniques for noninvasive monitoring of hemorrhage" Journal of Trauma, 71(1), 25-32, 2011. doi: 10.1097/TA.0b013e3182211601

[21] R. Fusco, M. Sansone, S. Filice, G. Carone, D. M. Amato, C. Sansone, A. Petrillo, "Pattern recognition approaches for breast cancer DCE-MRI classification: A systemic review", Journal of Medical and Biological Engineering, 36(4), 449-459, 2016. DOI: 10.1007/s40846-016-0163-7

[22] A. Huong, X. Ngu, "Neural network approach for rapid prediction of transcutaneous oxygen saturation", in 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE), Malaysia, 2019. DOI: 10.1109/ISCAIE.2019.8743751

[23] H. Y. Ma, Z. Zhou, S. Wu, Y. L. Wan, P. H. Tsui, "A computer aided diagnosis scheme for detection of fatty liver in vivo based on ultrasound kurtosis imaging" Journal of Medical Systems, 40(33), 1-9, 2016. DOI: 10.1007/s10916-015-0395-z

[24] A. Doronin, I. Meglinski, "The application of a unified Monte Carlo model in the training of artificial neural networks for the purpose of real-time in-vivo sensing of tissue optical properties" in 2019 SPIE Defense and Commercial Sensing, Baltimore, Maryland, United States, 2019. https://doi.org/10.1117/12.2518566

[25] A. Roggan, M. Friebel, K. Dorschel, A. Hahn, G. Muller, "Optical properties of circulating human blood in the wavelength range 400 – 2500 nm", Journal of Biomedical Optics, 4(1), 36-46, 1999. https://doi.org/10.1117/1.429919

[26] P. Mitra, C. A. Murthy, S. K. Pal, "Density-based multiscale data condensation" IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(6), 734-747, 2002.

[27] Y. Masuda, Y. Ogura, Y. Inagaki, T. Yasui, Y. Aizu, "Analysis of the influence of collagen fibres in the dermis on skin optical reflectance by Monte Carlo simulation in a nine-layered skin model", Skin Res Technol. 24(2), 1-8. 2017. https://doi.org/10.1111/srt.12421

[28] M. N. S. Yusiff, M. S. Jaafar, "Effect of multidesign skin model and characteristic on Monte Carlo simulation of light skin diffuse reflectance spectra", Jurnal Teknologi, 78(9), 135-142, 2016. DOI: 10.11113/jt.v78.7605

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 340-345 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Design of Efficient Convolutional Neural Module Based on An Improved Module

Daihui Li*, Chengxu Ma, Shangyou Zeng

Guangxi Normal University, College of Electronic Engineering, 541000, China

ARTICLE INFO

Article history:
Received: 05 November, 2019
Accepted: 19 January, 2020
Online: 07 February, 2020

Keywords:
Deep learning
Convolutional neural network
Feature extraction

ABSTRACT

In order to further improve the feature extraction performance of the convolutional neural networks, we focus on the selection and reorganization of key features and suspect that simple changes in the pooling layers can cause changes in the performance of neural networks. According to the conjecture, we design a funnel convolution module, which can filter out the key features and perform multi-scale convolution of key features. And we apply this module to the design of high performance small neural networks. The experiments are carried out on 101_food and caltech-256 benchmark datasets. Experiments show that the module has higher performance than classical module, and the small convolution networks based on the module has less parameters and more excellent performance.

1. Introduction

In 2012, deep learning shined on the ImageNet large scale visual recognition competition (ILSVRC), and more and more researchers began to notice deep learning. As an important branch of artificial intelligence, deep learning has been widely used in information processing in the fields of speech, image and text. Convolutional neural networks are the most classical network structures in the field of image processing, and they has excellent spatial feature learning abilities. The classical method to improve the convolution neural networks is to deepen and broaden the convolution neural networks, but it is easy to cause a large increase in the number of neurons. Large increases in the convolutional neural networks of parameters can lead to over-fitting, so the performance of neural networks will decline. Therefore, the model design of convolutional neural networks is particularly important. The excellent algorithm model can greatly improve the recognition accuracy and effectively reduce the computational complexity of the model, which determines the performance of almost all deep learning algorithms. However, the neural networks has so far been regarded as a black box device. In many cases, neural networks are not interpretative, and their design and improvement only rely on experience.

In this paper, we have improved the Reticulated Convolution Module (RCM) [1] and designed a small high performance convolutional neural network (HPCNet) with less parameters. Networks with fewer parameters are more easily ported to devices with limited transmission bandwidth and storage, such as cloud services, mobile platforms and smart robots. In addition, in the process of improved RCM, the pooling layer of the convolutional neural networks have attracted our attention. Simple changes in the use of different dimensional features lead to significant performance gaps in convolutional neural networks. This difference is instructive for the interpretability of convolutional neural networks and the design rules of convolutional neural networks. This paper mainly includes three parts of work: 1. We improved RCM and named it the second version of Reticulated Convolution Module (RCM-V2). 2. We designed a small convolutional neural network structure called backbone network (BBNet). And based on RCM-V2 and BBNet, a small convolutional neural network with high performance (HPCNet) was designed. 3. Finally, We evaluated RCMNet-V2 and HPCNet performance on two typical image classification datasets.

2. Related Work

Because of the birth of AlexNet [2], the design of convolutional neural networks has become an important research direction for scholars in the field of machine vision. The network structure of AlexNet has eight layers, including five convolution layers and three full connection layers. AlexNet adopts data enhancement methods and adds a dropout layer to prevent over-fitting. And it won the championship with a recognition accuracy of 10.9% ahead of its nearest rival. Since then, ILSVRC competitions had produced a large number of excellent network structures. VGGNet [3] explores the correlation between deep network and performance by continuously stacking small convolution kernels and pooling layers. GoogLeNet [4] transforms full connection and even general convolution into sparse connection to solve the defect,

* Daihui Li, Email: ldhcv@sina.com

but it keeps dropout. Two auxiliary softmax are added to the networks for forward conduction gradient to prevent the gradient from disappearing. GoogLeNet innovatively proposes the Inception structure which gathers the features of different scales together, expands the network width and enhances the expressive power of the networks. However, the deeper the network is, the more likely it is to cause degradation. In order to solve this problem, ResNet [5] came out. It introduced identity mapping into network structure and surpassed human in top-5 precision. ResNet is a milestone network, which makes it possible to train deeper networks. DenseNet [6] was born on CVPR2017 which established the dense connection between the front layers and the back layers. The connection of features between channels enables the network to realize high-density feature reuse, which achieves the performance beyond ResNet under the condition of less parameters and calculation cost.

With the increasing scale of the networks, resource asymmetry and other problems are caused in most scenarios. Therefore, compression and acceleration of algorithm models under the guarantee of network performance is a hot research topic in the field of network structure optimization. Lin M et al. [7] first proposed 1x1 convolution kernel in networks in order to greatly lessen network parameters. Iandola, F N et al. [8] proposed a lightweight network called SqeezeNet which got the AlexNet-level accuracy. Howard A G et al. [9] put forward the depthwise separable revolution in MobileNet. The depthwise separable revolution greatly reduces the parameters when the precision is slightly lower than the traditional convolution. However, the depthwise separable convolution reduces the information integration performance of the networks to a certain extent. Later, the idea of ResNext [10] was introduced into ShuffleNet [11] to clean the channels of convolution neural networks. Shufflenet strengthens the information circulation between channels, and improves the information expression ability of networks. The iconic NAS [12] promoted the spring tide of automatic machine learning (AutoML), which searched neural network structure according to reinforcement learning such as ENAs [13], DARTS [14] and NAO [15], etc. These models generated by automatic machine learning are not universal. It is also inconvenient to deploy an automated machine learning framework on some simple tasks, so designing an efficient small-convolution neural network is especially important. This paper aims to solve the problems of model design in simple machine vision tasks, and proposes an efficient micro convolution neural network called HPCNet. At the same time, some design inspirations of convolutional neural networks are obtained, which will provide inspiration for the design work of convolutional neural networks in the future.

3. Design of Small High Performance Convolutional Neural Network

This chapter first introduces the group convolution and the group convolution used in RCM. Then we introduce the design of RCM-V2 and its key features extraction process. Finally, the overall architecture of BBNet is introduced.

3.1. Group Convolution

Group convolution plays a crucial role in the design of RCM and RCM-V2. As shown in Figure 1, the group convolution groups the input features and then performs convolution operations within each group. When the size of the input feature maps is H*W*C and the output quantity of model is N, these feature maps are to be divided into G groups. After grouping convolution operation, each set of feature maps is C/G, and the output is N/G. It can be seen that the group convolution can consolidate information in small groups, which functions as feature isolation and cross interaction, and reduces the amount of parameters.

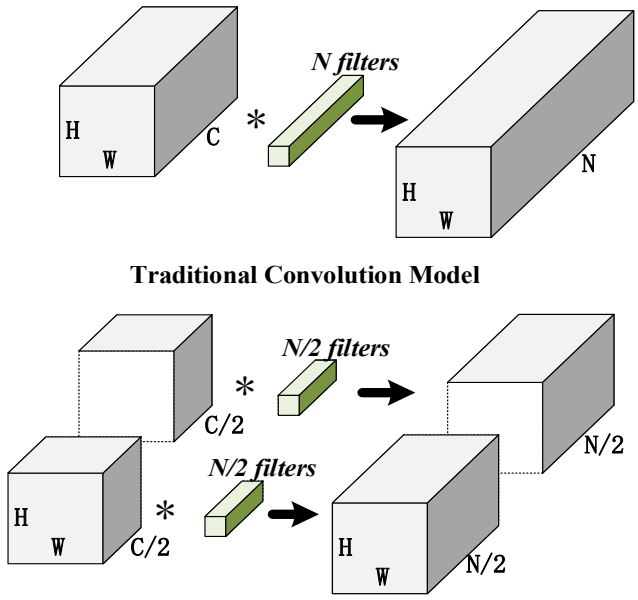


Figure 1. Group Convolution Model and Traditional Convolution Model

If the input of the convolution layer is composed of two different sets of features, and the number of feature maps in these two groups is not equal, great effect will be produced. As shown in Figure 2, A and B are two different sets of features, and the number of feature maps contained in A and B is different. A contains 128 feature maps and B contains 64 feature maps. Group convolution concatenates A and B, and then divides A and B into C and D. Both C and D contain 96 feature maps, and the feature maps in C are all derived from A. In D, 32 feature maps are derived from A, and 64 feature maps are derived from B. The convolution operation æ performs feature extraction on C, which is equivalent to feature selection and fusion on the first 96 convolution maps of A. The convolution operation £ performs feature extraction on D, which is equivalent to the feature fusion of the 32 feature maps of A and all the feature maps of B. As a result, F aggregates the features of A and B, while E merely extracts the features of A. Therefore, the features extracted by E and F will be significantly different.

3.2. Design of RCM-V2

The traditional convolution neural networks are formed by alternately stacking convolution layers and pooling layers. The desired effect is achieved based on the design of the kernel size of convolution layers and the design of different pooling layers. Single channel and single scale convolution settings tend to make the feature extraction of neural networks inadequate and constrain the performance of neural networks. In this paper, the overall structure of the mesh convolution module is funnel-shaped mesh.

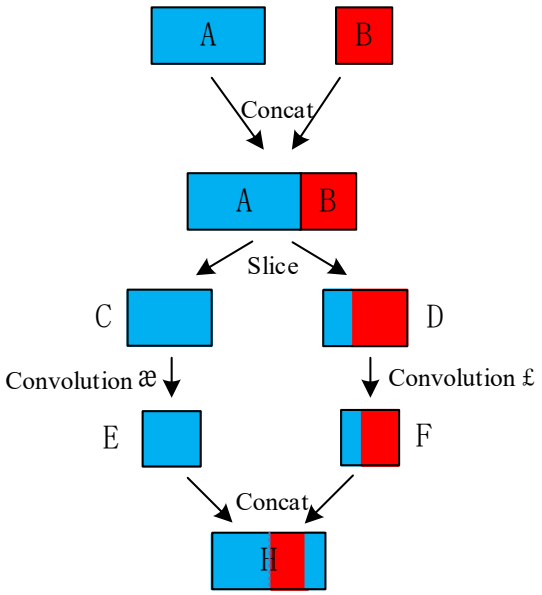


Figure 2. Group convolution of different feature sources.

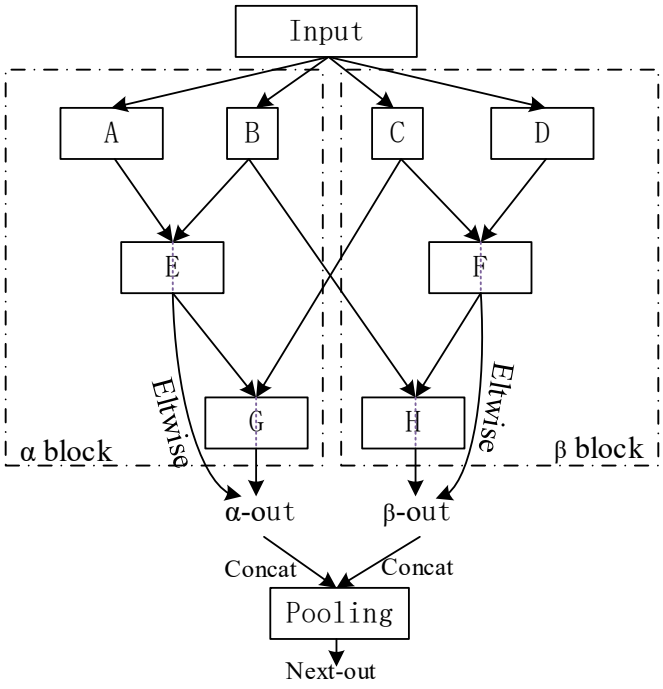


Figure 3. The structure of the RCM.

The overall structure of RCM is symmetrical, and the number of convolution kernels is unbalanced. RCMNet first integrates different features, and then combines them through cross structure. It not only realizes feature diversification, but also reuses a large number of important features, which makes convolution neural effectively enhance the ability of extracting key features. For miniature machine vision tasks, RCM can extract key features efficiently and ensure the diversity of features, so that convolution kernel can be used efficiently. Figure 3. shows the overall architecture of RCM. RCM is divided into α blocks and β blocks, and the key features are output of the bottom layer. In general, max-pooling preserves the main features. We focus on the pooling layers to further strengthen the feature selection ability of RCM.

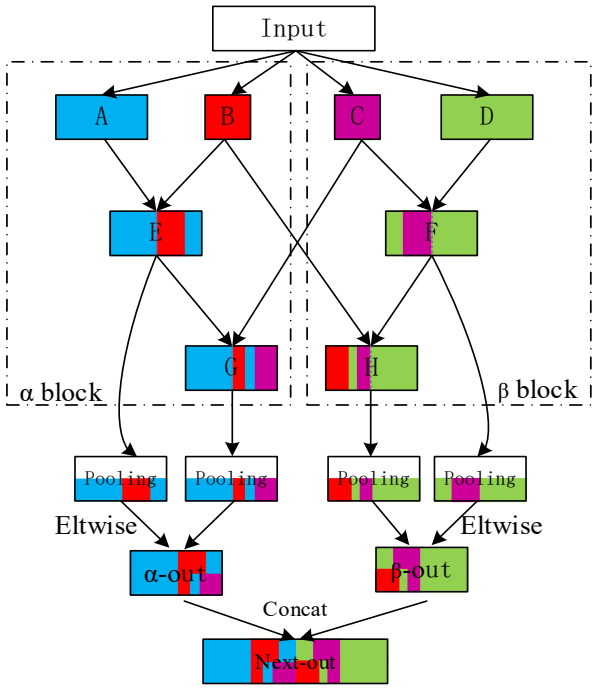


Figure 4. The structure of the RCM-V2. The number of groups for E, F, G and H layers is 2. In addition to the stitching method indicated in the figure, the remaining stitching method is Concat.

RCM-V2 has redesigned the connection between the RCM tail and the pooling layers. The structure of RCM-V2 is shown in Figure 4. The interlacing of different colors indicates the fusion of different features. When the input features enter RCM-V2, they first go through four convolutional layers called A, B, C, and D. They perform four different combinations of input features. B and C are convolution layers with a small number of convolution kernels, which can gather input features well. Next, the feature maps enter four interlaced grouping convolution layers which be named E, F, G and H. These four grouping convolutions interleave and fuse the key features of red and purple, which are like scattering water on the lotus' leaves and then aggregating into water polo. The reuse of key features effectively reduces the number of redundant features. The pooling layers of RCM-V2 directly extract the maximum value of receptive field from E, F, G and H, which effectively avoid the offset of features. Finally, half of the output features of RCM-V2 are multi-scale fusion results of multi-source features, and half are the results of extracting simple convolution features. Furthermore, RCM-V2 can be equivalent to a set of dual channel convolutions and a set of complex multi-scale convolutions.

Table 1: Parameters of RCM-V2

type	Layer	filter size/stride /output numbers/pad
Standard convolution	--	3*3/1/M/1
RCM-V2	A	1x1/1/ (1/2)*M/0
	B	1x1/1/ (1/4)*M/0
	C	1x1/1/ (1/4)*M/0
	D	1x1/1/ (1/2)*M/0
	E	3x3/1/ (1/2)*M/1
	F	3x3/1/ (1/2)*M/1
	G	3x3/1/ (1/2)*M/1
	H	3x3/1/ (1/2)*M/1

[a.] *M is the number of output features of the convolutional layers*

When the number of the input and output of the model's features is the same, the parameters of RCM are only equivalent to 91.67% of the conventional convolutional layers. The parameters of RCM-V2 are shown in Table 1.

This paper focuses on the pooling layers to improve the RCM. When the original RCM is connected to the next pooling layer, the characteristics of different dimensions are first eltwise and then input into the pooling layers. Although the original way can retain the features of each dimension to the greatest extent, it weakens the main features to a certain degree. In this paper, the pooling layers are advanced, and the method of post eltwise is used, and max-pooling is used to strengthen ability of expression the main feature of a single convolution layer. The expression of the original RCM is (1).

$$\begin{cases} \alpha_{out} = O_G + O_E \\ \beta_{out} = O_H + O_F \\ RCM_{out} = g(\alpha_{out} + \beta_{out}) \\ Next_{out} = p(RCM_{out}) \end{cases} \quad (1)$$

Where O_I is the output of layer I, α_{out} is the output of α block, βout is the output of β block, RCM_{out} is the output of RCM, $g(\cdot)$ represents concat operation, and $p(\cdot)$ represents max-pooling operation. $Next_{out}$ is the output of the RCM that connects max-pooling to the next layers. The expression of RCM-V2 is (2).

$$\begin{cases} \alpha_{out} = p(O_E) + p(O_G) \\ \beta_{out} = p(O_H) + p(O_F) \\ RCM_{out} = g(\alpha_{out} + \beta_{out}) \\ Next_{out} = RCM_{out} \end{cases} \quad (2)$$

RCM-V2 cleverly applies the max-pooling layers, which reduces the feature cancellation effect of adjacent dimensions to a certain extent, and strengthens the main features extracted by the core convolutional layers.

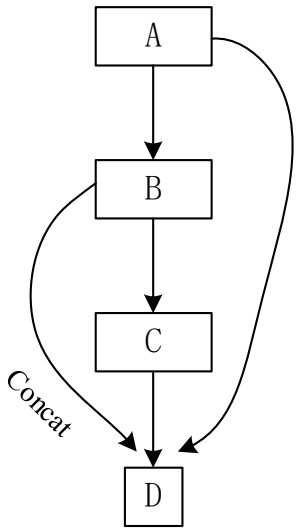


Figure 5. Inspiration of BBNet

3.3. Design of BBNet and HPCNet

When people understand or classify new things, they usually divide them into a large category, and then summarize the small classification to which the new things belong. BBNet refers to human cognitive mechanism of images, integrates high-dimensional features of different levels, and effectively uses high-dimensional features. This feature reuse mechanism of BBNet can use fewer convolution kernels for extracting features effectively and avoid extracting repeated features. To a certain extent, it can reduce the parameters of convolution neural networks. Specifically, the tail of BBNet uses cancat layers to splice high-dimensional features together, and then uses 1x1 convolution to check these features for information reorganization. Furthermore, in order to ensure the universality of BBNet, BBNet adopts a unified number of convolution kernels. Figure 5. shows the inspiration of BBNet and the Figure 6. shows the structure of BBNet.

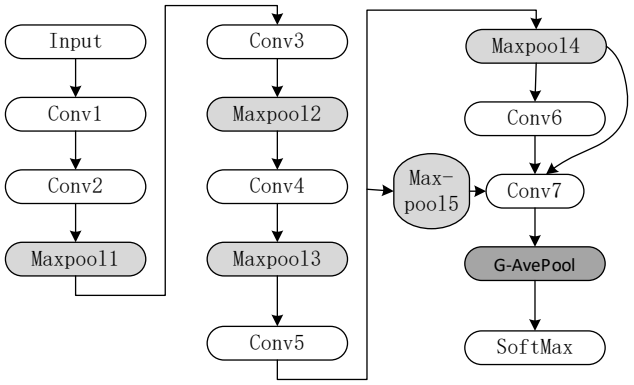


Figure 6. Structure of BBNet

We replaced the convolution module except Conv1 and Conv7 with RCM-V2 to build HPCNet. Table 2. shows the parameter Setting of HPCNet.

Table 2: Parameter Setting of HPCNet

Type	output size	filter size /stride/pad
Input	227*227*3	--
Conv1	75*75*96	7x7/3/1
RCM-V2	75*75*96	3x3/1/1
Maxpool1	37*37*96	3x3/2/0
RCM-V2	37*37*96	3x3/1/1
Maxpool2	18*18*96	3x3/2/0
RCM-V2	18*18*96	3x3/1/1
Maxpool3	9*9*96	3x3/2/0
RCM-V2	9*9*96	3x3/1/1
Maxpool4	4*4*96	3x3/2/0
RCM-V2	4*4*96	3x3/1/1
Conv7	4*4*101/--	1x1/1/0
G-AvePool	1*1*101/--	--
SoftMax	1*1*101/--	--

4. Experiments and Results Analysis

In order to better evaluate the performance of RCM-V2 and BBNet, this paper conducted experiments in two datasets, 101_food and Caltech-256. The experiment was divided into two parts. The first part was to compare RCM-V2 with the original RCM, and the second part was to test BBNet and HPCNet.

4.1. Experimental Platform and Data Processing

Experimental platform: The experiment uses caffe as a deep learning framework, and uses Ubuntu 16.04 as the operating

system. The computer has 32GB of RAM, a four-core and eight-thread i7-6700K CPU and a NVIDIA GTX-1080Ti GPU.

Datasets: 101-food contains 101 kinds of food and each kind of food contains 1000 pictures. It contains a total of 101000 pictures. The ratio of training set and test set is 3:1. Caltech-256 includes 256 categories. The number of pictures in each category is not fixed. There are 29781 pictures in total. In this paper, the training set and test set are divided into 4:1. 101_food is a large and medium-sized food classification data set, which has high requirements for fine-grained image recognition of neural networks. Caltech-256 is a kind of data set with many kinds, but it needs convolution neural networks with strong generalization ability because of the uneven distribution of pictures between classes.

Data pre-processing: The entire pre-processing process can be divided into three steps. Firstly, the images of all data sets are unified to the size of 256 * 256 for scale normalization. Secondly, we need to enhance the data. We choose to randomly cut 227 * 227 sub graphs from 5 directions of all the pictures (left upper corner, top right-hand corner, left lower corner, bottom right-hand corner and the middle), and turn the sub graphs horizontally to achieve the effect of expanding the original data set tenfold. Finally, the mean value is normalized to ensure that the mean value of all features is near zero.

4.2. Results and Analysis

The number of images of each category in 101_food is equal, and a dataset with a uniform number of images is very suitable for network benchmarking. Caltech-256 is a dataset containing many categories, but the images of each category are unevenly distributed. It requires that the networks have higher generalization performance and can extracting features batter. We evaluate RCMNet and RCMNet-V2 on these two datasets. Figure 7, Figure 8. and Table 3. show the accuracy curve of RCMNet and RCMNet-V2 on different datasets.

Table 3: Performance of RCMNet and RCMNet-V2 on 101_food, Caltech-256

Network	Model Size(M)	Top-1 Acc (%)	Datasets
RCMNet	2.1	69.55	101_food
RCMNet-V2	2.1	70.35	
RCMNet	2.2	59.05	Caltech-256
RCMNet-V2	2.2	59.81	

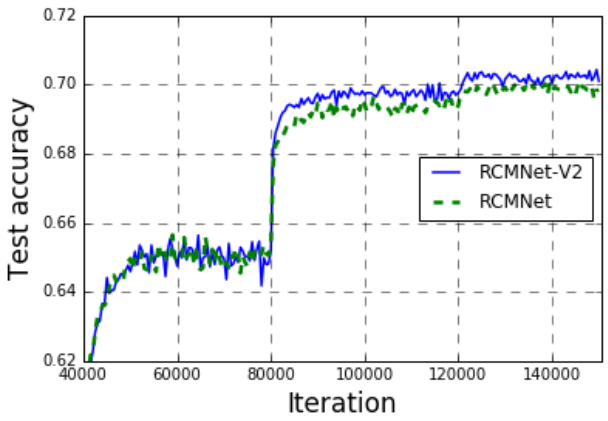


Figure 7. The accuracy of RCMNet and RCMNet-V2 on 101_food

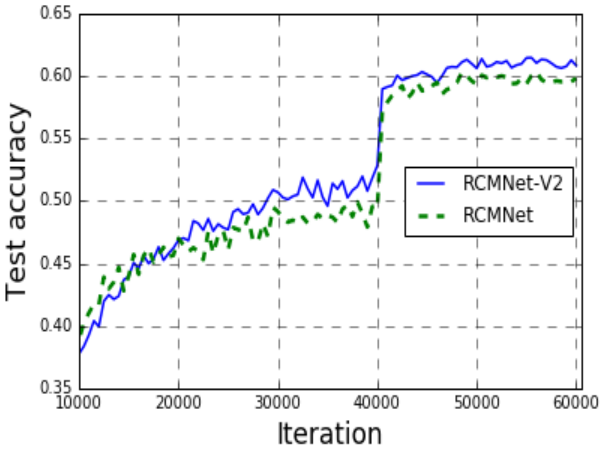


Figure 8. The accuracy of RCMNet and RCMNet-V2 on Caltech-256

RCMNet-V2 has a certain performance improvement on both datasets, which proves the effectiveness of RCM-V2. Compared to RCM, RCM-V2 only changes the connection of the max-pooling layers. The improved version of the RCM-V2 has the same amount of parameters as the RCM and does not increase the amount of calculation.

Next, we will evaluate the performance of BBNet and HPCNet. Similarly, we perform classification tests on these two datasets. Table 4 shows the evaluation results.

Table 4: Performance of BBNet and HPCNet on 101_food, Caltech-256

Network	Model Size(M)	Top-1 Acc (%)	Datasets
BBNet	1.8	63.58	101_food
HPCNet	1.7	69.85	
BBNet	1.9	54.96	Caltech-256
HPCNet	1.8	60.10	

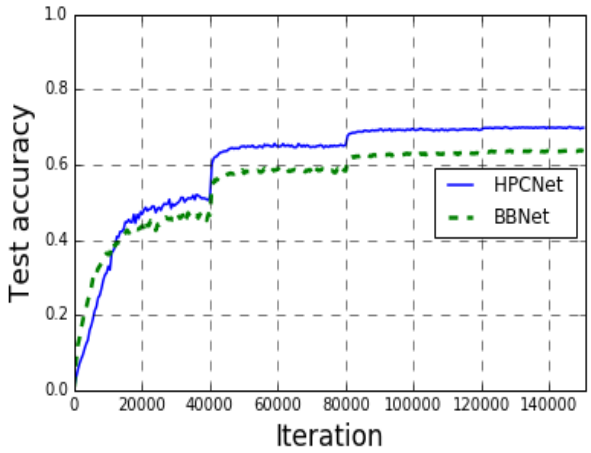


Figure 9. The accuracy of BBNet and HPCNet on 101_food

As can be seen from Figure 9 and Figure 10, BBNet has a higher accuracy rate than HPCNet at the beginning of training, but in the middle and late training period, HPCNet's accuracy is significantly higher than BBNet. It has proven that HPCNet with integrated RCM-V2 has a higher performance improvement than BBNet. Next, We compare HPCNet with the traditional networks.

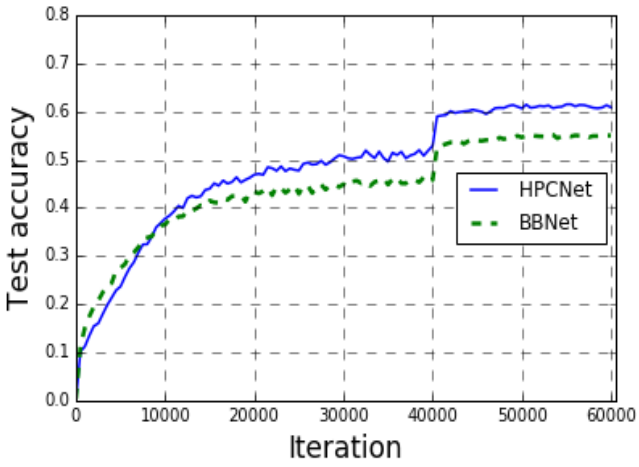


Figure 10. The accuracy of BBNet and HPCNet on Caltech-256

Table 5: Performance of BBNet and HPCNet on 101_food, Caltech-256

Network	Model Size(M)	Top-1 Acc (%)	Datasets
SqueezeNet[8]	3.2	56.3	
AlexNet[2]	234.8	56.7	
VGG[16]	553.6	59.3	101_food
ResNet[16]	95.1	67.4	
HPCNet	1.7	69.8	
AlexNet[2]	236.8	56.0	Caltech-256
HPCNet	1.8	60.1	

Table 5 shows that HPCNet has smaller sizes and it has higher accuracy than popular networks. Even with one of the classic compression networks called SqueezeNet, HPCNet still has higher accuracy and smaller sizes. Moreover, HPCNet's equal convolution kernels design makes it highly modifiable. For different tasks, the number of convolution kernels can be changed to achieve better results.

5. Conclution

This paper focuses on the max-pooling layers to improve the performance of RCM and proposes RCM-V2. Experiments have shown that direct pooling that be used in our paper can get good effect for RCM. A simple change of the application mode of pooling can bring about a great improvement in performance. It brings some inspiration to the design of convolutional neural networks. For example, when should we fuse the features rather than strengthen them directly? All in all, the solution of this problem can help us to design convolutional neural networks quickly and understand convolutional neural networks better.

A small high-performance convolutional neural network named HPCNet is designed and its performance is evaluated on two datasets of different types. HPCNet has higher performance and smaller sizes than the classic networks on two classic datasets.

Conflict of Interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in the work we submitted.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (grant no. 11465004).

References

[1] Daihui Li, Shangyou Zeng. Chengxu Ma, “Design of Convolution Neural Network Based on Reticulated Convolution Module” in 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 2019. https://doi.org/10.1109/ICEIEC.2019.8784673

[2] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks” Communications of The ACM, 60(6) 84–90, 2017. https://doi.org/10.1145/3065386

[3] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015. https://doi.org/ 10.1109/cvpr.2015.7298594

[4] Simonyan, Karen , and A. Zisserman . "Very Deep Convolutional Networks for Large-Scale Image Recognition." Computer Science, 2014.

[5] He, Kaiming, et al. "Deep Residual Learning for Image Recognition." in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. https://doi.org/ 10.1109/cvpr.2016.90

[6] G. Huang, Z. Liu, L. van der Maaten, and K. Q. "Weinberger. Densely connected convolutional networks" in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. https://doi.org/10.1109/cvpr.2017.243

[7] Lin, Min, Qiang Chen, and Shuicheng Yan. "Network In Network." Computer Science, 2013.

[8] Iandola, Forrest N. , et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size" arXiv Preprint arXiv:1602.07360, 2017.

[9] Howard, Andrew G. , et al. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications" arXiv Preprint arXiv:1704.04861, 2017.

[10] S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, "Aggregated Residual Transformations for Deep Neural Networks" 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017. https://doi.org/ 10.1109/cvpr.2017.634

[11] X. Zhang, X. Zhou, M. Lin and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices" 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018. https://doi.org/ 10.1109/cvpr.2018.00716

[12] Zoph B , Vasudevan V , Shlens J , et al. "Learning Transferable Architectures for Scalable Image Recognition" in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. https://doi.org/10.1109/cvpr.2018.00907

[13] Pham H , Guan M Y , Zoph B , et al. "Efficient Neural Architecture Search via Parameter Sharing" arXiv Preprint arXiv:1802.03268, 2018.

[14] A. Ahmed et al., “Modeling and Simulation of Office Desk Illumination Using ZEMAX,” in 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 2019, pp. 1–6.

[15] Luo R, Tian F, Qin T, et al. "Neural Architecture Optimization" arXiv Preprint arXiv:1808.07233, 2018.

[16] Heravi, Elnaz Jahani, Hamed Habibi Aghdam, and Domenec Puig. "An Optimized Convolutional Neural Network with Bottleneck and Spatial Pyramid Pooling Layers for Classification of Foods." Pattern Recognition Letters, 105, 50–58, 2017. https://doi.org/10.1016/j.patrec.2017.12.007

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 346-354 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Project-Based Learning as a Tool of Enhancing of Entrepreneurial Attitude of Students

Dana Paľová[*,1], Martin Vejačka[1], Lukáš Kakalejčík[2]

[1]*Department of Applied Mathematics and Business Informatics, Faculty of Economics, Technical University of Košice, 04001, Slovakia*

[2] *GrowthPro, Košice, 04001 Slovakia*

ARTICLE INFO

Article history:
Received: 06 December, 2019
Accepted: 13 January, 2020
Online: 07 February, 2020

Keywords:
Entrepreneurial education
Project-Based Learning
Virtual laboratory
Europe 2020

ABSTRACT

The increase in employment and productivity is crucial for the economic growth of the European Union. Specifically, the employment of young people needs to be increased. The excessively theoretical approach to education is identified as one of the most significant problems in the educational system. Acquirement of practical experience while study, presents for the students the advantage when applying for a job or starting their own business. Self-employment or entrepreneurship might be one of the partial solutions for high youth unemployment in EU countries. The information and communication technologies transform education processes already for decades. The education in the area of entrepreneurship is also the field where these transformations occur. New educational forms and tools are being used in this area. Project-Based Learning and virtual laboratory are two examples of such new forms of education supported by information and communication technologies used in entrepreneurial education. This paper aims to present how project-based learning and virtual laboratory can support entrepreneurial education on the examples of two courses offered at our faculty. Through project-based learning, participants of these courses achieve the opportunity to implement their knowledge, skills, and experience to solve practical problems. It, all together, enhances the students' motivation and involvement in the educational process. They also practical experience establishing and managing a small virtual business within the virtual laboratory of entrepreneurial education. The paper describes the setup of these courses, their content, and evaluation by students and achieved results. The findings show that project-based learning and the virtual lab supporting entrepreneurship education via electronic tools are useful ways of entrepreneurial skills development.

1. Introduction

This paper is an extension of work originally presented in MIPRO 2019 [1]. The creation of new businesses always supported the economic growth of the economy [2]. Education in the field of entrepreneurship and enhancing the entrepreneurial attitude of students should be an important part of the curricula of universities with majors in economics and business management. However, in Slovakia, universities in the economics study plans usually prepare their students for the role of employees not for the role of entrepreneurs.

Harris [3] adduced that education often does not provide practical experience with real-world entrepreneurship. On the other hand, the time during the studies at the university could be an ideal time to start-up student's business, while they have the opportunity to practically learn, cooperate with or employ their classmates allowing them to gain work experience. At the same time, the support of the student's family in the case of business failure can be counted as one of the biggest benefits. Furthermore, a lot of successful businesses occur after multiple business failures [4]. In addition, failed enterprise brings valuable experience for involved entrepreneurs, which can be used in their future careers (both as entrepreneurs or employees). Information and communication technologies (ICT) influenced the world of business in many ways. It is currently common, that ICTs are used

[*] Dana Paľová, Faculty of Economics, Technical University of Košice, Nemcovej 32, 04001, Slovakia, +421556023268, dana.palova@tuke.sk

for setting up the businesses, for promoting and distributing products, for using the banking services or for communication with authorities. The impact of ICT was also reflected in the emergence and use of electronic commerce (or e-commerce) as a still-growing distribution and marketing channel [5].

Therefore, entrepreneurship courses set up and provided at our university aim to provide the opportunity to gain practical experience with entrepreneurship for our students.

2. Theoretical Background

2.1. Entrepreneurship and unemployment of the young generation in the context of European Union growth and sustainability

Youth unemployment is one of the key economic problems in multiple EU economies. In 2018 the youth unemployment was over 15 percent in the European Union, which is higher in comparison with other productive groups of citizens [6]. One of the possible solutions on how to reduce the youth unemployment rate might be the support of their self-employment ambitions by fostering the entrepreneurial spirit in the education process. Entrepreneurship supports personal development, encourages active participation in a community or society, and teaches a person how to enter or reenter the job market as a self-employed person or an employee.

The Europe 2020 strategy urges to promote economic growth and competitiveness via innovation [7]. Entrepreneurship in innovative fields might help to achieve its goals. The individual's ability to turn ideas into action, the activity aimed at achieving the business goals and the implementation of innovation and creativity are the main characteristics of the entrepreneurship itself [8]. New Skills Agenda for Europe [9] promotes the development of the entrepreneurial skills in the Entrepreneurship Competence Framework (Entrecomp) [10]. EntreComp defines entrepreneurship as an ability necessary in the 21st century, including social, managerial, technical, cognitive and relational skills. McCorkle et al. [11] also indicate these skills as important for entrepreneurs. Wiklund and Dean [12] include also creativity as a crucial entrepreneurial skill. Table 1 (bellow) illustrates EntreComp Framework and its main parts visualization.

Table 1: EntreComp Framework

Key Area	Competency
Ideas and Opportunities	• Spotting Opportunities • Creativity • Vision • Valuing Ideas • Ethical and Sustainable Thinking
Resources	• Motivation and Perseverance • Self Awareness and Self Efficacy • Financial and Economic Literacy • Mobilizing Others • Mobilizing Resources
Into Action	• Learning Through Experience • Working with Others • Planning and Management • Taking the Initiative • Coping with Ambiguity, Uncertainty and Risk

Source: [9]

2.2. Education towards innovative entrepreneurship

When starting a new business, entrepreneurial skills are very important. To know how to reduce business risks, to know how to use the creativity tools, to know how to learn from the market and to know-how to innovate are the key competencies in the realization of a new business idea [13]. Some of the main entrepreneurial skills for successful entrepreneurship were described by the Institute of Entrepreneurship Development [14]:

- Creating a business plan and planning in general together with the access to financing are crucial factors for business sustainability.
- Communication and networking skills to clearly express ideas and create relationships with people within and outside the organization.
- Leadership and managerial skills covering skills for the ability to plan organizational activities, coordination and motivation of human resources and developing cooperation and team-work, communicating the vision of the business and its objectives, etc.
- Digital marketing skills to boost competitiveness, innovation, and productivity. These skills include the know-how in the usage of digital tools and technologies for business growth on the market.

2.3. Project-Based Learning method (PBL) in Education of 21st Century

The current generation of students was born into a world of ubiquitous Information and Communication Technologies (ICT). Therefore, it is natural for them to use ICT in every aspect of their lives, including education. Brown [15] stated that for digital youth, ICTs are an integral part of their social life, which changes their perception of life, behavior, and education. Also, their requirements on the educational process, design and content are influenced by their attitude towards ICT. Millennials spend on average 6.5 hours using ICT for different purposes (e.g. listen to and record music; view, create and publish Internet content; play video games; watch television; talk on mobile phones and instant message) [16]. They can be described by the following attributes: group-oriented, global, technologically confident, risk-taking and optimistic. Taking into account the above-mentioned facts, the International Education Advisory Board [17] defined main requirements on new learning design as following (Table 2):

Table 2: Requirements on 21st-century learning design

Requirements	Learning design
Well-defined goals	Learning objectives are clear – the teacher must define in advance clear, realistic, reachable conditions for the successful accomplishment of the learning process. The main idea is that everybody has the same criteria for passing, it doesn't matter on the way, speed and time needed to complete the objectives.
Patience	Students may redo assignments repeatedly if necessary. It is important to give a chance to try it more times until the students are finally successful. The Millenials are prepared to learn from their mistakes and they will try to do the task again and again.
Team play	Tasks are completed, evaluated and reviewed by the groups of students. Using this way, the learners achieve the possibility to introduce themselves, to identify their strong and weak characteristics and discover their uniqueness and individuality.
Tracking	Students could see their progress at any time and any particular education process level and

Requirements	Learning design
	also as a whole. It allows them to control their path to reach the learning objectives.
Change	Students are learning time management and how to progress towards (achieving) the goal by dividing the large tasks. At the same time, they can learn how the task division moves a project along.
Immediate consequences	Teachers provide immediate/useful feedback and meaningful, formative and diagnostic assessments. It helps students to identify, what works for them and what does not in an easier way.
Personalization	Students can discover their strengths and reach the learning objectives by playing different roles.
Patterns	The learning environment makes sense for the students. By learning to identify patterns and make logical decisions during the education process the environment helps students cultivate higher-level thinking skills.

Source: [17]

The project-based learning (PBL) might be one of the methods suitable for teaching this new generation of digitally native students. The Gold Standard of project-based learning (Figure 1) introduced by [18] includes also other skills such as self-reflection, critical thinking, research and inquiry skills.

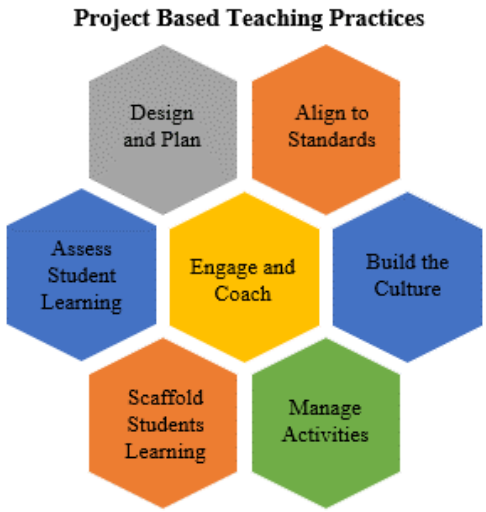


Figure 1: Key Knowledge, understanding and success skills enhanced by Gold Standard PBL; source: [18]

Project-Based Learning includes the personalization of learning for students [19]. PBL is student-directed learning that requires students to work on real-world projects over a specific period. During this time, they have to present and implement different knowledge and skills. Therefore, students find out that obtained skills and knowledge can be used in the practice [20].

While using PBL methods, lecturers could create collaborative groups of students to encourage the learning of students, as they have to face the same difficulties together [21]. The learning process then includes the utilization of the prior knowledge and information and promotes discussion in the students' groups. This leads to a higher interest of students in the educational process [22].

While using Project-Based Learning, the project solved should meet the following points:

- *Authenticity* – the project should be an authentic problem solved by using real tools, processes, and standards. The main principle is, that project needs to speak to personal concerns, interests, and issues of particular student.
- *Challenging Problem or Question* - makes learning more meaningful for students as they have to apply a lot of their knowledge and skills to find a solution to the problem.
- *Sustained Inquiry* - as students' colleagues or lecturers constantly confront their achieved results, the searching for necessary information is more dynamic and interactive. This inquiry might include interviews with experts, future customers, etc.
- *Student Voice and Choice* – students receive only project definition, but the methods of the project's working-out depend on students' choice.
- *Reflection* – supports what students have learned and encourages them to imagine how it could be applied elsewhere - outside the project. Teacher together with the students reflects the quality of student work, obstacles that arise and together try to find the problem overcoming strategies.
- *Revision* and *Critique* – show how to receive (and also give) constructive feedback during the project development (thanks to that students get experience how it runs in real life). Experts from practice or classmates might be involved in a critique.
- *Public Product* – increases motivation to higher quality work of students, because they are afraid of possible negative feedback on their results' presentation before the audience consisting of their classmates [18].

Project-Based Learning might support creativity amongst the students and teachers. Furthermore, students deepen the context of their knowledge and develop communication skills and critical thinking as they are doing an authentic project [23].

2.4. Meaning of the virtual laboratory in modern education

Currently, one of the crucial factors influencing the competitiveness and survival of the business on the marketplace is the rate of effective exploitation of ICT. The entrepreneurs must be able to keep track of a huge amount of information and available tools and to determine its possible implementation to their business area. This fact points to the need to integrate ICT into the entrepreneurial education process. The use of information and communication technologies in education may have many forms and various intensity.

The use of a virtual laboratory for entrepreneurial education is rather new. The intensive ICT involvement in teaching and learning process could be represented by the virtual laboratory. The concept of the virtual laboratory could be described in general as an environment of various cooperating ICT systems enabling the simulation of some real-world activity [24]. The usage of the virtual laboratory in entrepreneurship education might provide real-market experience for students in a very cost-effective way. Klamma et al. stated that a virtual laboratory might be very effective in ICT usage in entrepreneurial education [24].

Typically, the virtual laboratory is considered to be an interactive environment for carrying out simulated experiments.

The laboratory can be characterized as "a playground for experimentation" providing tools that can be used to manipulate objects relevant to a specific scientific domain. This concept is relevant for both, conventional and virtual laboratories. However, there is an important distinction in terminology, where a virtual laboratory is often custom-built to perform a limited set of experiments [25].

Task of design and development of the virtual laboratory for the education is highly complex. It covers the need of the set of the skills from various areas such as pedagogy, design and visualization. This process involves also the development of texts, images, learning environments and their interactions [25]. The development of the virtual laboratory and it's implementing in the teaching process, builds up the knowledge in technology, pedagogy and content knowledge [26].

Bretz et al. [27] indicated that involving virtual laboratories to the education process brings many advantages. It helps students to understand deeper science concepts, to develop their ability to identify, evaluate and solve different kinds of problems [28]. Moreover using virtual laboratories increases students' curiosity and grows up their motivation to take part in the education actively [29].

Virtual laboratories are becoming an increasingly common tool for teaching [30]. However, this is true, especially for the fields of physics, chemistry, and biology [31]. It is very often used for entrepreneurial education because it has significant potential in this area [24]. Often, the virtual laboratory in the field of entrepreneurial education is implemented differently than in the natural sciences, where it frequently represents virtual reality simulation of a lab. When used in entrepreneurial education, the virtual laboratory more likely simulates the functionality of various parts of the economy, a firm or a firm's lifecycle. The examples of simulated objects might be business registration or interaction with the public administration authorities, payment systems, market mechanisms, online marketing tools, and many others. This brings great possibilities of gaining valuable experience for the students without the necessity of great expenses spent on running a real business and involving virtual laboratory in that field supports meaningful learning by linking the previous information with a new one [32].

3. Courses supporting Entrepreneurial Education

All startups face great challenges when trying to grip on the market and up to 90% will fail to survive [33]. Da Silva et al. [34] adduced pricing issues, market validation and ignorance of customers' demands as main problems causing startup failures. Wang et al. [35] confirmed the fitting of the products on the market as an important problem of the startups adding also the functionality of the business model and issues with the acquisition of customers. Giardino et al. [36] also included targeting too-specialized markets to the list of significant issues of startups causing their failure. In consideration of these facts, the education process of the future entrepreneurs should help them to prevent such issues.

Furthermore, mastering ICT usage is currently a crucial element of entrepreneurial success. Entrepreneurial education should reflect the reality and the usage of ICTs as the most innovative and beneficial for students. McCorkle et al. [11] stated that students studying marketing and entrepreneurial education need space for creativity, what traditional education usually limits. The overview of Granitz and Pitt [37] summarized several unusual technology tools to teach entrepreneurship and marketing.

Resulting from the research in entrepreneurial and marketing education carried out, two new university courses were introduced. The courses should allow to apply students' knowledge in entrepreneurship (and in marketing as its organic element) and acquire practical skills in real conditions with the help of modern ICT tools. The above mentioned PBL method widely supported by various ICT tools was used to develop the content of the courses.

The education towards entrepreneurship is more-less covered by most of the subjects learned at the Faculty of Economics. During their studies, students are educated in theoretical fields like management, company finance, accounting, legislation, etc. This theoretical knowledge is important to achieve, but it cannot be denoted as the engine of students' curiosity and self-involvement in their study process. That is why it is really important to provide subjects that provide them more practical insight into the "real" world of the business and entrepreneurship. This is the reason why we developed innovative courses "Marketing Tools for the Presentation and Evaluation of Business Ideas" and "Economic Information Systems". In the mentioned courses, students are led to go through almost all levels of the entrepreneurial life cycle described in [38] and at the same time, they develop their skills defined in EntreComp, which are important for start-uppers and new entrepreneurs in general.

3.1. The course: Marketing tools for the presentation and evaluation of business ideas

The first course named "Marketing tools for the presentation and evaluation of business ideas" aims to concede practical education in the area of first steps in the entrepreneurial life cycle, for example, the introduction of new idea or product, and afterward company/corporate marketing. Within this course, students solve the practical problems of developing the product, fitting the product for the market, promotion, and analytics. This allows students to experience the start of the company from scratch widely supported by various ICT tools [1].

Figure 2: Final web-site of the service supporting sale of used cars; source: own based on students' web-site

The content of the course could be described as following: students should bring the business idea of product creation and then directly verify its market readiness and response. Students gain theoretical knowledge in the area of business models, key performance indicators, business objectives, web design, online marketing and its data analysis. When solving the practical problems, students obtain practical experience with social network marketing and applications such as Google Ads, Google Analytics,

Google Trends, Google Tag Manager, Google Market Finder, WordPress, and others. To achieve the final goal, i.e. to make a decision about the competitiveness and sustainability of their product or service, students need to design a web page, which is used for collecting main data for the Google Analytics tool. Figure 2 and Figure 3 represents the preview of designed websites.

Figure 3: Final web-site of the service supporting the free call of a new service provider; source: own based on students' web-site

The course has two parts: face-to-face (F2F) part, where its content is dedicated to basic project management, various business models and their Key Performance Indicators, multiple types of marketing (email, social media, content marketing), web design, marketing research, and marketing data analysis and self-study part, where the LMS Moodle is used for publishing and management of the course materials, schedules, and activities. The acquired skills and knowledge are then used by the students during work on their projects. Students form small teams (2-5 people) and their roles in the team are assigned by themselves. The student teams present their business model, product, product's fit to market and marketing strategy. The team projects are evaluated by lecturers, people from outside (operating in the field of online marketing and web-page design) and other students with a focus on business model definition and objectives, web presentation, usage of the marketing tools, gathering of the relevant data and its analysis and product viability on the market.

3.2. The course: Economic Information Systems

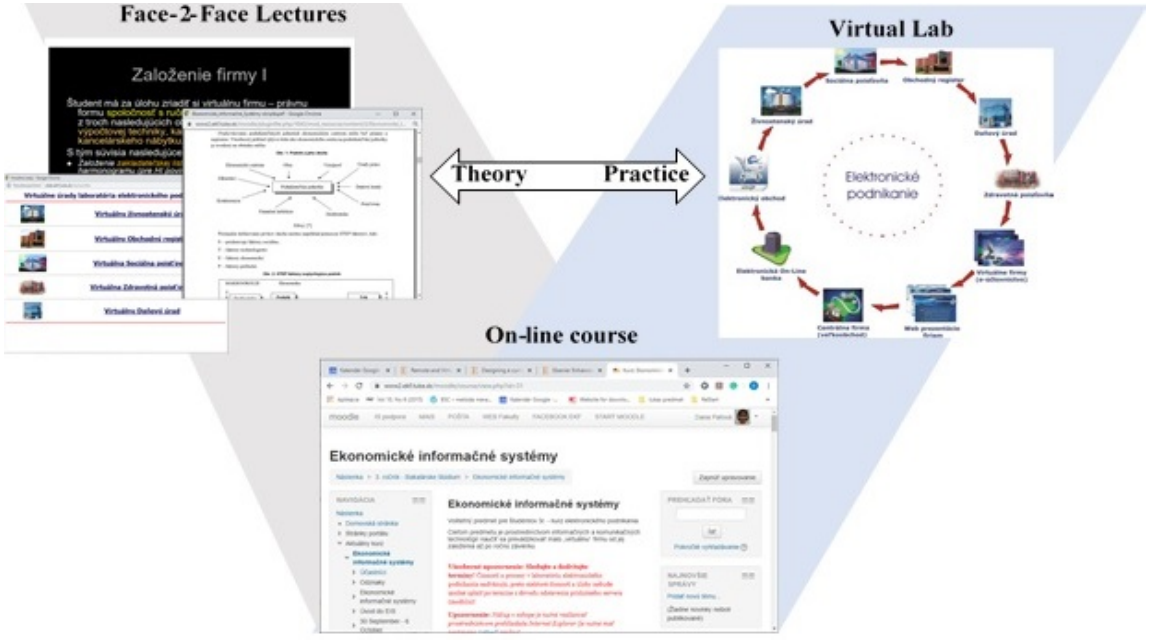
The provision of products using information and communication technologies is in the world of the business closely bounded to business intelligence systems, interchange of electronic data or electronic payments. All of these are known as electronic commerce and it can provide important tools and services for different business' relations as business-to-business, business-to-customer or customer-to-customer relations [39].

From the point of view of entrepreneurial education, it seems to be important to present the main activities needed by the entrepreneurs in the area of electronic commerce, for example, communication with the public administration authorities to meet the legal obligations, use of e-banking and e-payments [40]. The above-mentioned activities, important during the first period of the enterprise's existence, represent the content of the second developed course "Economic Information System".

The main idea of this course could be described in Figure 4.

The Virtual Lab itself is a part of e-course, which is providing students with various study materials, for example, information on how to manage a virtual firm from its establishment to the end of the first business year using ICT. The on-line course, operated by LMS Moodle, acts as the main source where all relevant information is collected. Students can find there most important dates and the description of the tasks needed to be done.

The virtual laboratory covers the following parts of electronic entrepreneurship: authorities of the business registration (providing social and health insurance registration, taxation, licensing, etc.), virtual e-shops (both retail and wholesale), virtual bank and other entities necessary during firms' life-cycle. Students establish their business after getting familiar with all the legislative requirements, therefore they need to follow the legislation of the Slovak Republic and prepare all necessary documents for establishing the business (e.g. deed of foundation, registration forms, licenses, etc.).



Afterward, they have to operate their own business using company website, e-shop. During the subject study, they act also as clients of businesses of their classmates. Using this principle, we can simulate the real market and students could achieve experience in pricing, sales and payments coordination. At the same time, students need to keep an accounting of their business and prepare the financial statements, which are bases for the final strategic study of their business. The results of the particular part of the course are evaluated by university staff – experts in particular fields, for example, accounting, legislation, information technologies, etc. To be able to provide all content of the course fully using e-way, different tools are involved within the virtual lab, such as EJBCA (the open-source PKI certification authority), Adobe Acrobat Reader DC ©, UNICOM Electronic Online Bank [5].

4. Results and discussion

The main objective of both described courses is to provide practical insight into the company's life-cycle to the students. Every course shows the different parts of the entrepreneurship. While the course "Marketing Tools for the Presentation and Evaluation of Business Ideas" covers the beginnings of the business – just from the idea appearance to a decision whether the product could survive on the marketplace, the course "Economic Information Systems" presents the process from real establishment of the company till first year anniversary. To be able to evaluate, whether our goal was achieved, both of the courses contain the evaluation of the students' skills and knowledge acquired and also the feedback, which is used to evaluate the content of the courses and way how they were provided.

Answering the electronic feedback questionnaire was last mandatory task for all course participants in both courses. Respondents (using the five-point Likert scale) should indicate the degree of agreement with the group of statements aimed at the quality and realization of particular parts of education process in the course. Gathered rates of agreement were presented in the form of percentages of agreement.

The course of "Marketing Tools for the Presentation and Evaluation of Business Ideas" is optional, what causes the number of enrolled students is usually lower than in the mandatory courses. The feedback and evaluation of students are gathered in each course iteration.

Students acquire the final evaluation of their performance at the end of the course with the maximum percentage of 100% in the total. Following table contains the results of students during the recent iteration of the course (Table 3):

Table 3: Mean values reached by students for specific parts of the projects

The part of the project evaluated:	Maximum value	Mean value
Formulation of objectives	10%	8.5%
Web design and its functionality	15%	13.5%
Web analytics	15%	14%
Content marketing	15%	11.5%
Data Analysis	10%	9.75%
Viability of product	10%	10%
Presentation of the project	25%	25%

Source: own contribution

At least one member of each team had to achieve official Google certification in Google Analytics Fundamentals and all students' teams have to achieve Google Tag Manager Fundamentals certifications. More than half of the students participating in the course successfully achieved these certificates. The teams often struggled to clearly define business goals and to set up Google Analytics, which lead to the misinterpretation or insufficiency of collected data from online channels.

Furthermore, students had to answer the feedback questionnaire at the course's end. The summary of the questionnaire's results can be following:

- 53% strongly agreed and 47% of students agreed that the content of the course met their expectations.
- 54% of responding students stated as the most interesting the part of designing the web and its further development, the marketing part of the course was indicated by13% of them and 33% did not have the favorite part of the course.
- Over 87% of students were satisfied with the educational style of the course, mainly they highlighted not obvious borderline between theoretical lectures and practical exercises. Overall, students evaluated the course as motivating and interesting.
- 87% of the respondents indicated the course as moderately difficult and 13% as very difficult.
- All students agreed that the acquired skills and knowledge during the course are applicable in entrepreneurial practice.
- All participants appreciated the necessity to use teamwork and creativity in the process of their course project development.
- All of the responding students confirmed that the completion of the course improved their understanding of the operation of the start-up.
- 93% of all students managed to assess the importance of the business idea.
- 87% of respondents indicated the improvement of their managerial and entrepreneurial skills.

During the EIS course, students can obtain a maximum evaluation of their performance at the level of 100% at the end of the course. The following table contains their results during the recent iteration of the course (Table 4):

Table 4: Mean values reached by students for specific parts of the course

The part of the course evaluated:	Maximum value	Mean value reached
Establishing a company according to legislation requirements	20%	18.13%
Business plan preparation	15%	12.58
Website preparation	10%	9.42%
Eshop administration	15%	13.55%
Accounting	20%	16.47%
Strategic study	10%	8.75%
Analysis of the information system used in the course	10%	9.12%

Source: own contribution

During each year of Electronic Information Systems course teaching, surveys to gather feedback on the course are conducted. The survey is conducted after the course completion in each curse

iteration. Investigated participants' opinions on the EIS course are divided between the parts such as starting-up the business, e-payments, communication with public administration authorities and e-commerce. The students' answers also provide suggestions for improvement of the education process in the course. The survey completion was mandatory for all participants of the course. In the most recent survey, 117 completed questionnaires were gathered.

The main results of the recently conducted survey can be summarized as follows:

- Almost 95% of students reported, that they acquired significant experience with establishing and management of the start-up company during the EIS course.
- Over 85% percent of respondents adduced gaining new information and experience with the process of the business establishment in accordance with the valid local legislation.
- Over 98% percent of participants agreed that obtaining practical experience with e-communication with virtual public administration authorities using the electronic registry was instructive.
- Almost 95% of the respondents highlighted the use of the digital signature in electronic communication with public administration authorities as valuable experience.
- Above 42% of the course participants indicated that the creation of the business plan of their virtual company was beneficial for them. All of the students already had experience with writing the business plan during previous years of their studies.
- More than 29% of students adduced that creating their company's website company was beneficial, even though they already had created a website before in other courses.
- Setting-up and managing their firm's e-shop was a significant experience for more than 78% of participants. Mainly, the operation of an e-shop from the administrator's side was highlighted by the participants.
- Almost 63% of students appreciated the possibility to try out to keep an accounting of their firms. Especially, the experience in the first business period accounting was emphasized.
- Approximately 65% of respondents also welcomed the preparation of a strategic plan of their firms for the upcoming years (beyond the horizon of the ongoing course).
- Most of the course participants (93%) evaluated the whole course of Electronic Information Systems very positively in general. Moreover, they appreciated the opportunity to practically experience the entrepreneurial processes within the virtual laboratory of electronic entrepreneurship.

In general, it can be concluded, that both described courses "Marketing Tools for the Presentation and Evaluation of Business Ideas" and "Economic Information Systems" stimulated the higher activity of the students. Furthermore, the course participants also often discussed the various issues and topics during the course lessons in order to solve the problems with projects. In spite of short time of the course duration (three months), students handled the product development, the creation of the website, to preparation and execution of marketing campaign and analysis of the data collected (as for the "Marketing Tools for the Presentation and Evaluation of Business Ideas" course) and experience the first year of existence of virtual start-up company (as for the "Economic Information Systems" course).

In both cases, students welcomed this opportunity and highly evaluated the course's benefits. The usage of ICT tools and virtual laboratory for entrepreneurial education appears to be beneficial according to our results. Also, the participants' practical experience with the projects is among the most important outcomes of both courses.

In connection with the EntreComp Framework, students achieved competencies from the parts of "Ideas & Opportunities" (especially vision, creativity, evaluation of ideas and sustainable thinking), "Resources" (financial and economic literacy, mobilizing resources) and "Into Action" (learning through experience, working with others, planning and management, coping with ambiguity, uncertainty, and risk) [10].

5. Conclusion

Despite different actions supporting the entrepreneurship from the EU like "Funding opportunities" [41], "Erasmus for Young Entrepreneurs" [42] or "EU Startup Services" [43], it is not trivial for young people to start their own business. The business sector is challenging and changing at a rapid pace and it is hard to enter this world without any previous knowledge and skills.

In [44], 34% of self-employed respondents stated the need to take an entrepreneurship course before starting up their own business, because they perceive an insufficiency of experience and skills necessary to ensure the sustainability of the business in the labor market. At the same time, from this research follows, that just less than half of EU citizens perceive that their education gave them the know-how and the skills to enable them to run a business (41%) or allowed them to understand the role of entrepreneurs in society (47%) [44].

This was the motivation to innovate courses in a way to implement most of the missing entrepreneurial skills into the education process.

The course of "Economic Information Systems" allows participants to acquire practical entrepreneurial experience in the field of e-commerce using the virtual laboratory of electronic entrepreneurship. Students gained practical experience with setting-up business according to the Slovak Republic legislation. Moreover, they experienced the usage of digital signature and the electronic registry in communication, the creation of the business plan, the preparation of the company's website with implemented e-shop and its administration. Finally, the course participants tried to keep accounting and to create a strategic plan for the future periods of their virtual firms. The use of the virtual laboratory for the purposes of entrepreneurial education is unique in educational processes of Slovak universities.

The course of "Marketing Tools for the Presentation and Evaluation of Business Ideas" allows students to experience practical education from the introduction of a new product and its marketing. Within the course, its participants solve the problems of developing the product, product's fit to the market, marketing analytics, which allows them to experience the launch of the

business supported by ICT tools. The students have to come with a business idea and confirm its viability. Furthermore, they gain theoretical and practical knowledge in business models, online marketing, web design, and data analysis. Moreover, the course participants have to design a web page, which is also used for collecting data for Google Analytics. Students also gain practical experience with network marketing, WordPress, Google Ads, Google Trends, Google Tag Manager, Google Market Finder.

In both courses, students highly evaluated the courses' benefits and the practical experience gained during the participation in the courses. By graduating described courses, participants got a chance to achieve or improve their knowledge and skills listed in EntreComp Framework. Step by step, by taking part in the time and solving the semestral projects, they built up the particular competencies needed in the business sector, such as the ability to be innovative and creative and define the problems and find suitable solutions. The usage of virtual laboratory and Project-Based Learning for entrepreneurial education appears to be beneficial according to results acquired.

Conjunction of these courses create educational synergies for better preparation of students for their possible entrepreneurial future. Passing these courses increases chances of successful self-employment of courses' graduates or even the creation of new work-positions for the others as they might start-up new businesses in real economic environment. It also might lower the unemployment rate in given economy, what is closely connected to the objectives stated in Europe 2020 strategy [7].

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This contribution was supported by the national project "Decision Support Systems and Business Intelligence within Network Economy" (Contract No. 1/0201/19) funded by Grant Agency for Science; Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

[1] L. Kakalejčík, D. Paľová, "Enhancement of students' skills via project-based learning" in MIPRO 2019: 42nd international convention. Croatian Society for Information and Communication Technology, Electronics and Microelectronics, 769-774, Rijeka, Croatia, 2019. http://dx.doi.org/10.23919/MIPRO.2019.8756958

[2] J. Gerdin, G. Linton, "Contingency fit(s) in entrepreneurship research: uses and usability", Örebro University School of Business, 2016.

[3] K. Harris, Education and Knowledge: the structured misrepresentation of reality, Routledge, 2016.

[4] P. Thiel, B. Masters, Zero to One: Notes on Startups, or How to Build the Future, New York: Crown Business, 2014.

[5] M. Vejačka, "Electronic commerce entrepreneurship education using virtual laboratory" in MIPRO 2019: 42nd international convention. Croatian Society for Information and Communication Technology, Electronics and Microelectronics, 670-675, Rijeka, Croatia, 2019. http://dx.doi.org/10.23919/MIPRO.2019.8757070.

[6] Statista, "Youth unemployment rate in EU member states as of December 2018", Statista, 2018.

[7] European Commission, "EUROPE2020: A European strategy for smart, sustainable and inclusive growth", European Commission, 2010.

[8] Council of the European Union, "Council conclusions on entrepreneurship in education and training", Education, Youth, Culture and Sport Council meeting, Council of the European Union, 2014.

[9] European Commission, "Analytical Highlight Focus On Entrepreneurial skills", EU Skills Panorama, European Commission, 2015.

[10] European Commission, "EntreComp: The Entrepreneurship Competence Framework", JRC Science for Policy Report, European Commission, 2016.

[11] D. E. McCorkle, J. M. Payan, J. Reardon, N. D. Kling, "Perceptions and reality: Creativity in the marketing classroom" Journal of Marketing Education, SAGE Publications, Thousand Oaks, Canada, 29(3), 254-261, 2007. http://dx.doi.org/10.1177/0273475307306892

[12] .J. Wiklund, S. Dean, "Entrepreneurial Orientation and Small Business Performance: a Configurational Approach" Journal of Business Venturing, 20, 71-91, 2005. http://dx.doi.org/10.1016/j.jbusvent.2004.01.001

[13] .G. T. Lumpkin, G. G. Dess, "Linking two Dimensions of Entrepreneurial Orientation to Firm Performance: The Moderating Role of Environment and Industry Life Cycle" Journal of Business Venturing, 16, 429-451, 2001.

[14] Institute of Entrepreneurship Development, "4 Important Skills for Young Entrepreneurs", Institute of Entrepreneurship Development, 2017.

[15] J. S. Brown, "Learning in the Digital Age" The Internet & the University: Forum 2001, 65-91. 2001.

[16] D.F. Roberts, U.G. Foehr, V.Rideout, "Generation M: Media in the Lives of 8-18 Year-olds", A Kaiser Family Foundation Study, 2005.

[17] International Education Advisory Board, "Learning in the 21st Century: Teaching Today's Students on Their Terms", International Education Advisory Board, 2007.

[18] Buck Institute for Education, "Gold Standard PBL: Essential Project Design Elements", PBL works, Buck Institute for Education, 2018.

[19] S. Boss, "Project-Based Learning: A Short History", Edutopia, 2011.

[20] Buck Institute for Education, "What is PBL?", PBL works, Buck Institute for Education, 2019.

[21] C. E. Hmelo-Silver, "Problem-based learning: What and how do students learn?" Educational Psychology Review, 16(3), 235–266, 2004. http://dx.doi.org/10.1023/B:EDPR.0000034022.16470.f3

[22] H. G. Schmidt, S. M. M. Loyens, T. Van Gog, F. Paas, "Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006)" Educational Psychologist 42(2), 91–97, 2007.

[23] M. J. Sousa, M. do R. Almeida, "Entrepreneurial Skills Development" in Recent Advances in Applied Economics, Lisbon, Portugal, 2014.

[24] R. Klamma, M. Spaniol, D. Renzel, "Virtual entrepreneurship lab 2.0: sharing entrepreneurial knowledge by non-linear story-telling" Journal of Universal Knowledge, 1(3), 174-198, 2006.

[25] B. S. Wästberg, T. Eriksson, G. Karlsson, M. Sunnerstam, M. Axelsson, M. Billger, "Design considerations for virtual laboratories: A comparative study of two virtual laboratories for learning about gas solubility and colour appearance" Education and Information Technologies, 24(3), 2059–2080, 2019.

[26] M.. Koehler, "What is technological pedagogical content knowledge (TPACK)?" Journal of Education, 193(3), 13-19, 2013. http://dx.doi.org/10.1177/002205741319300303

[27] S. L. Bretz, M. Fay, L .B. Bruck, M. H. Towns, "What faculty interviews reveal about meaningful learning in the undergraduate chemistry laboratory" Journal of Chemical Education, 90(3), 281-288, 2013. http://dx.doi.org/10.1021/ed300384r

[28] L. B. Bruck, M. Towns, S. L. Bretz, "Faculty perspectives of undergraduate chemistry laboratory: goals and obstacles to success" Journal of Chemical Education, 87(12), 1416-1424, 2010. http://dx.doi.org/10.1021/ed900002d

[29] A. H., Johnstone, A. Al-Shuaili, "Learning in the laboratory; some thoughts from the literature" University Chemistry Education, 5(2), 42-51, 2001.

[30] D I. Lewis, "The pedagogical benefits and pitfalls of virtual tools for teaching and learning laboratory practices in the biological sciences" The Higher Education Academy: STEM, 2014.

[31] K. Achuthan, V. K. Kolil, S. Diwakar, "Using virtual laboratories in chemistry classrooms as interactive tools towards modifying alternate conceptions in molecular symmetry" Education and Information Technologies, 23, 2499–2515, 2018. https://doi.org/10.1007/s10639-018-9727-1.

[32] A. Hakim, A. Liliasari, A. Kadarohman, Y. M. Syah, "Effects of the natural product mini project laboratory on the students' conceptual understanding" Journal of Turkish Science Education (TUSED), 13(2), 27-36, 2016.

[33] N. Patel, "90% Of Startups Fail: Here's What You Need To Know About The 10%", Forbes, 2015.

[34] F. R. da Silva, R. Fabrício, P.R. da Silva, N.V. Galegale, G. K. Kabane, "Why technology-based startups fail? An IT management approach" in Production and Operations Management Society, POMS 26th Annual Conference, Washington, D.C., USA, 2015.

[35] X. Wang, H. Edison, S. S. Bajwa, C. Giardino, P. Abrahamsson, "Key challenges in software startups across life cycle stages" in International Conference on Agile Software Development, Springer International Publishing, 169, 2016. http://dx.doi.org/10.1007/978-3-319-33515-5_14

[36] C. Giardino, S. S. Bajwa, X. Wang, P. Abrahamsson., "Key challenges in early-stage software startups" in International Conference on Agile Software Development, Springer, Cham, Switzerland, 2015. http://dx.doi.org/10.1007/978-3-319-18612-2_5

[37] N. Granitz, P. Leyland, "Teaching about marketing and teaching marketing with innovative technology: Introduction to the special edition" Journal of Marketing Education, 33(2), 127-130, 2011. http://dx.doi.org/10.1177/0273475311410844

[38] R. W. Price, Roadmap To Entrepreneurial Success Powerful Strategies for Building a High-Profit Business, Global Entrepreneurship Institute, AMACOM Books, American Management Association, 2004.

[39] E. Turban, D. King, J.K. Lee, T.P. Liang, D.C. Turban, "Electronic commerce" Prentice Hall Press Upper Saddle, NJ, USA, 2015. http://dx.doi.org/10.1007/978-3-319-10091-3

[40] V. Venkatesh, J. Y. L. Thong, F. K. Y. Chan, P. J. H. Hu, "Managing citizens' uncertainty in e-government services: the mediating and moderating roles of transparency and trust" Information Systems Research, 27(1), 1–218, 2016. http://dx.doi.org/10.1287/isre.2015.0612

[41] European Commission, "Funding opportunities for small businesses", European Commission, 2015.

[42] European Commission, "Erasmus for Young Entrepreneurs", European Commission, 2019.

[43] European commission, "EU Startup Services", European Commission, 2017.

[44] European Commission, "Entrepreneurship in the EU and beyond", Eurobarometer, European Commission, 2010.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 355-360 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Analysis of Reducing Waste on Defective Rubber Products in the Magnetic Disc Drive Industry

Hernadewita[1], Evan Haviana[1,*], Mohd Nizam Ab Rahman[2], Hendra[3]

[1]*Master of Industrial Engineering, Postgraduate Program, University of Mercubuana, 11650, Indonesia*

[2]*Department of Mechanical and Materials Engineering, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, Malaysia*

[3]*Department of Mechanical Engineering, Faculty of Engineering, University of Tirtayasa, Banten, 42191, Indonesia*

ARTICLE INFO

Article history:
Received: 25 October, 2019
Accepted: 01 December, 2019
Online: 07 February, 2020

Keywords:
Root cause analysis
FMEA
Productivity

ABSTRACT

Productivity is increasing when less input was used in getting the same outcomes. Thus, the effectiveness of productive effort, especially in industry, as measured in terms of the rate of output per unit of input. As in derivative rubber industry which is producing the seal product for final assembly of Disc Drive. The function of seal is to protect the instrumentality within magnetic disc. Throughout production runs, there were number of issues occurred inflicting a poor quality or high defect rate, such as the embrace machines that perpetually breakdown or stop and creating a less productivity within the production. The data used for this analysis are the Rubber production for seal product on year 2017 till 2018. One of the seven QC tools, Ishikawa diagram, and along with Failure Modes and result Analysis (FMEA) were applied. The results of the analysis are the improvement program among the production space of Rubber product and extend the standard by the foremost highest defect waste at rubber production processes. There were tends to decide it into auto-deflashing. The improvement result showed that the product defects were seen to be reduced by 35% (from 11.7% into 7. 6%) for auto-deflashing issues.

1. Preliminary

The current progressively competitive in global economy, the viability of Associate in maintaining business might be affected to each product and services industries. They should be concerned in meeting customers' needs as quickly as possible and turn out quality product and services at reasonable costs. Any business is challenged to boost its performance to reply quickly and accurately to changes that occur within the market. Meanwhile, some typically products are unable to compete with their competitors, due to the highest cost on production as the affected of ineffective and economical production processes.

As in the productions of disc drive business, competition within the technology of constructing exhausting drives is obtaining robust, Western Digital (WD) joined of the Disc Drive (HDD) producers was reported to closed one of their factories in Petaling Jaya, Selangor, Malaysia. At the end of 2019, the works begin to establish Solid Storage Disk (SSD) company within the Penang area, Malaysia and prepared to start producing in the next few months [1].

In addition to the WD hardware company, called disk drive products, Seagate also says it will cut 6,500 workers globally. This is similar to 14% of the entire workers at this point. The announcement was quoted by Kompas Techno from The Wall Street Journal, [2] created by Seagate for a while when the Cupertino, California-based company said it would cut 1,600 positions in the company. Seagate had about 45,000 full-time employees before the job reduction program was announced. HDD makers, along with their suppliers, must use all of your capabilities again to develop and formulate appropriate production methods and systems that are tailored to their business aspirations so that they can win products that are effective, economical, and quality so that they can survive in HDD, business division. To respond to the challenges of increasingly competitive business, generating business must develop and formulate existing production methods and systems that will be tailored to the business desire to win productive and economic productivity.

* Evan Haviana, Email: evanhaviana@gmail.com

Problems in the production floor of the "Gasket Connector" product continue to be high product deficiencies as there is no optimal method or machine that makes the product inferior to the customer's specifications of the product and quality refers to handling the seal itself

2. Literature

Theoretically, Yamit's book reveals that the actual method of assembly is the process of modifying (transforming) a material or element (entry) into another product which includes a higher price or in the process of adding a price [3]. As illustrated by the Figure 1 below regarding the installation process;

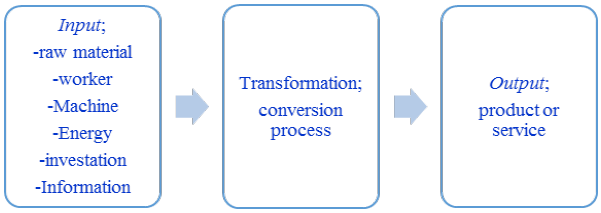


Figure 1: Production Process

Production planning and management involves planning activities with production activities so that planned activities are carried out well, starting with the important activities of raw materials, the amount needed after the product is finished and some resources required [4]. The most objective production planning is:

1) Maximizing service for purchasers,
2) Minimize investment in inventory,
3) Planning with production capability,
4) Production management,
5) Inventory,
6) Storage capability and
7) Instrumentality material movements, and
8) Routing and planning with processes.

1.1. Lean Manufacturing

Lean Production is a philosophical and practical way to eliminate all waste in the production process. Using lean concepts has practically improved the production process in a sustainable manner [5-6]. Lean Manufacturing is a good system used by most industries to identify and reduce waste through continuous improvement, it is done by streamlining products to attract customers to the highest possible [7-8].

1.2. Fishbone Diagram

There are many improvement strategies developed by many researchers along with the use of seven QC tools and FMEA [9]. One of them is a fishbone diagram, commonly known as the Diagram of the cause and the effect or the Ishikawa Diagram, introduced by Dr. Kaoru Ishikawa, a high quality professional management from Japan, collectively consists of seven basic quality tools (7 basic tools). Dr. Kaoru Ishikawa is Japanese, perhaps the World Health Organization's leading figures have introduced bone and fiber results to the world. The bone diagram (Figure 1) of Ishikawa has become a very widely used tool worldwide for causing issues. The rationale is simple, the bone diagrams make sense, and guide every team to consider finding the most appropriate explanation for drag. Bone analysis starts with the issues and the bones provide a template to separate and evaluate the reasons. There are usually six classes, but the quantity will vary depending on the matter. This technique allows for issues to be analyzed and when used with colleagues, it offers everyone an idea of the matter so that the solution can be developed together). Organizations where staff are inspired to evaluate practices, risks, and errors after they occur tend to have a culture in which root or bone analysis is used. It helps to understand the cause and to explain it[10].

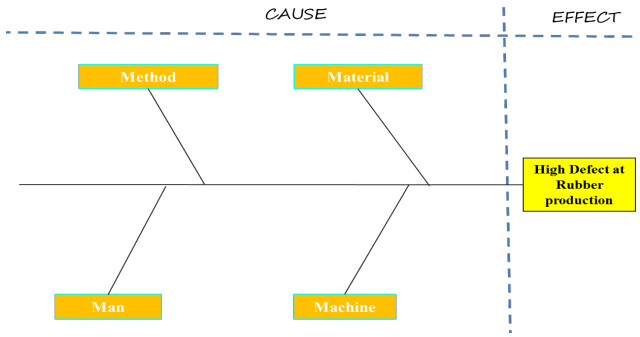


Figure 2: Typical Fishbone Analysis Diagram

1.3. Root cause analysis

Root cause analysis is increasingly being used in health and social services to improve safety and quality as well as minimize side effects as it provides a retrospective review of incidents or events [11]. Ishikawa teaches us to look inside by asking why? Just ask why we can find the real cause of the problem several times. The real reason is not to show. Using this Ishikawa diagram, it is easier for us to find the root cause of the problem, the manufacturing industry in which the process is identified with a variety of variables that can cause problems. If problems and causes are known as expected, corrective actions and steps are easier. With this diagram, everything becomes clearer and allows us to see all the possible causes and find out the real cause.

In performing Ishikawa Analysis, there are many steps to take, namely:
1. Prepare an Ishikawa analysis session.
2. Distinctive consequences or issues.
3. Establish numerous classes of main causes.
4. Finding potential causes by group action.
5. Review every class of main causes
6. Succeed agreement on the foremost possible causes

The strength of the Ishikawa diagrams / bone diagram is set up to illustrate any weaknesses and all involved will contribute to the suggestions that may be behind the problem. Disadvantages of the Ishikawa Diagram is an opinion-supported tool and designed to limit team / user flexibility in visualizing problems that soften to their liking. Further, because ballot paper is usually used to determine the main reasons listed in the diagram.

Ishikawa's diagram has been expanded into a cause and effect diagram. The development of the Ishikawa Diagram was performed using the "Why" technique up to 5 times (five reasons)

with known reasons behind the effects, it is hoped that the results of the assembly method are dynamically enhanced with the factors controlled by one method [12]. This figure is also useful in distinguishing potential causes of tangle. The subjective diagrams focus on the problem of accentuation or symptoms that are the main cause behind this problem. The causative diagram at the same time shows the cause of the problem by linking the causes together.

The advantage of Ishikawa diagram is that it will describe each downside that happens and everybody concerned in it will contribute recommendation which may be the reason behind the matter. Moderate weaknesses Ishikawa diagrams area unit opinion supported tools and are designed to limit the flexibility of the team/user visually in describing the matter victimization the deep "level why" technique unless the paper used is admittedly giant to suit those wants. Furthermore as typically ballot is employed to settle on the foremost possible cause listed on the diagram.

Root cause analysis and corrective action method are fully essential for the advance of the standard management system and increasing the standard of the ultimate product or service [13]. This text has intention to shortly highlight the foremost steps that ought to be taken within the right sequence so as to with success and for good resolve any downside from problematic method. Its 2 major areas as that are reciprocally reticulate and can't perform while not one another root cause analysis and corrective action method. The primary one serves the purpose to sight the proper root reason behind the matter that is that the supply of the problems and also the different one may be a set of actions to for good eliminate the root cause with the projected solutions that directly attack it so as to fully take away it from the method.

1.4. Failure Mode and Effect Analysis (FMEA)

FMEA is one in every of the tools of Six alphabetic character to spot the sources or causes of a high quality downside [14], in step with FMEA is done by:

- Identify and value the potential failure of a product and its effects.
- Identify actions which will eliminate or scale back the possibility of a possible failure to occur.
- Record method (documented the process).

1.5. Determination of FMEA Variable Value

The following are variables from FMEA, namely:

- Severity is graded in step with the seriousness of the potential impact of failure.
- The incidence is Associate in nursing acceptable rating to assess wherever the primary level causes and failure modes created by the corporate can occur.
- Detection (Detection) is graded in step with the likelihood that this sighting or management technique can detect a possible failure mode.

1.6. Risk Priority Number (RPN)

The Priority variety downside is articulated via RPN:

- RPN = Severity rating x incidence rating x Detection rating.
- Corrective action should be taken if the RPN is high.
- There is not any absolute rule for what's a high RPN variety. Conversely, FMEA is commonly seen as relative (the highest RPN is addressed first).

Which can be seen mathematically within the equation below.

$$RPN = Severity \times Occurrence \times Detection \quad (1)$$

1.7. The following are the Failure Mode and Effect Analysis Procedures

1. Create a table describing the values specified. For the Frequency of Occurrence, Degree of Severity, and Chance of Detection columns make a consensus table of relative values to assume occurrence, severity (how much effect the failure has on the effect), and the possibility that the problem is detected and resolved now this (detection). Fill in the appropriate value for the columns above based on the table made.
2. Calculate the risk factor for each cause of failure. For each cause of failure, the risk factor is the product of the numbers in the column in the Occurrence, Degree of Severity, and Chance of Detection columns.
3. Identify critical failure modes (has a large risk factor value).
4. Create columns in a spreadsheet. Give the names of each column as follows: Modes of Failure, Cause of Failure, Effect of Failure, Frequency of Occurrence, Degree of Severity, Chance of Detection, Risk Priority Number (RPN), and Rank.
5. Identify all possible modes of failure, can be done by brainstorming.
6. Identify all possible causes of failure for each mode of failure above.
7. Determine the effect of each failure. Identification of potential consequences of failure to customers, products, and processes.

3. Methodology

An analysis methodology can be a method or procedure that contains steps that are clearly and consistently set in the analysis method. Each and every level that determines the successive stages must be meticulously approved. In this study, the company that manufactures rubber for magnetic disc components at the Batam city area, the subject of the discuss in this paper is the rubber material for the seal products of the Rubber Department. As the Rubber Department is unable to meet the KPI (Key Performance Indicator) and this can be directly attributed to the waste/defect (PPM) contained in the Rubber Production Department. There are several stages of analysis to be carried out in this analysis. Among them, there were the theories learned and related to the theme of the research provided. These theories can then be compared to the knowledge gained from field studies or interviews with workers. Field studies was carried out in the form of studies conducted through direct observation in the production room, to obtain information related to installation methods.

4. Result and Discussion

The company is one in every of the businesses that support

magnetic disk drive parts within the world, settled in state. The main business is producing a part for magnetic disk Drive. The corporate was founded in 1993. Works professionally with ISO 9001, ISO 14001 and OHSAS 18001 standards. The company try to maintain of reasonable performance within the areas of Quality, environment, Safety, and Health. Currently, the companies for which this analysis is conducted are supported by more than 900 employees and mostly experts in the assembly room.

The selection of products that will be inspected as long as the company is based on a single family product, some of these products are considered one family if they are through regular means and use of public facilities. Within the product group, there are many products and therefore product choices are mapped based on many issues such as number of daily outputs, demand and frequency in the given volume. One of the methodologies used is the matrix integration of the method of assembly. The matrix of this production method can be a matrix containing all types of products. See the Table 1 below:

Table 1: Rubber Product Matrix

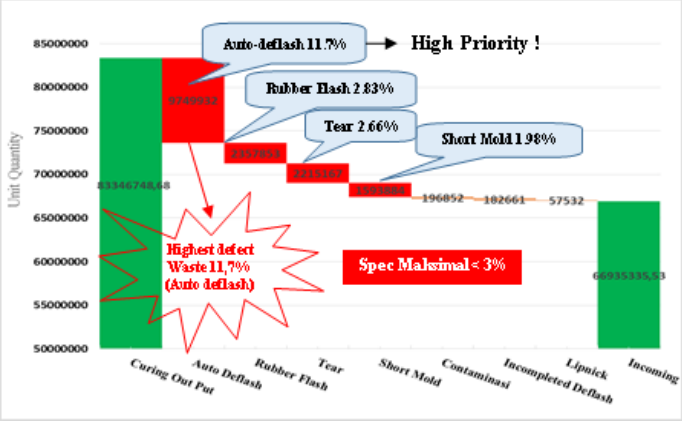


Table 2: Pareto data analysis for defect Product.

Product Category	Product Name	Picture	Process									
			Rubber Preparation	Curing	Post cure	Screening sheet	De-Flashing	Pre-Washing (Hand Rub & Jet Wash)	U/S washing	Oven Drying	Final Screening	Packing
DB00	Wider Lip		X	X	X	X	X	X	X	X	X	X
DB01	3,5" Seal Connector		X	X	X	X	X	X	X	X	X	X
DB02	2,5" Seal Connector		X	X	X	X	X	X	X	X	X	X
DB03	Connector Packing		X	X	X	X	X	X	X	X	X	X
DB04	Gasket Thin		X	X	X	X	X	X	X	X	X	X
DB05	Gasket Blade		X	X	X	X	X	X	X	X	X	X
DB06	Connector Packing		X	X	X	X	X	X	X	X	X	X
DB07	Connector Packing		X	X	X	X	X	X	X	X	X	X

Based on higher tables than Auto-deflash or self-deflash (11.7%), Flash Rubber (2.83%), Tear (2.66%) and Short Print (1.98%). then we can conclude that the symptoms of Auto Deflash Rubber are the highest defect and can be a top priority for us to conduct analysis and corrective action, which is also the main source of supply from the highest NG level (DPPM).) weaknesses in rubber products are eight families higher than and also the source of waste distribution due to product standards that do not meet specifications. In addition, this phenomenon is what specializes in project follow-up on the installation floor

4.1. Waste / Defective name recognition

Auto deflash on gasket products can be a product that cannot persist with the molding (Self-release from rubber sheet) once the molding method is complete and therefore the auto deflash product can cause breakage on parts/products, as shown in Figure 3 below:

Figure 3: *Defect auto Deflash*

Auto deflash leaves a tear problem on the product.

4.2. General Production Process Flow

General descriptions of the rubber method include rubber, curing, deflashing, and several other inspection and cleaning methods to support the method until it is an ambient process, as well as the seal and preparation before being sent to the client..

4.3. Analysis of Fishbone Diagrams

The fishbone diagram or cause and effect diagrams was applied in additional to analyze by input factors, by using 4M analysis (Man, Method, Materials and Machine), as shown on Figure 4 below:

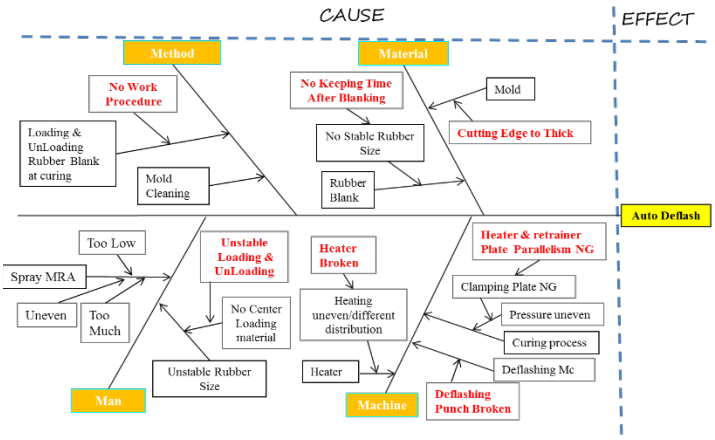


Figure 4: Fishbone diagram

As a result, from our investigation, authors were able to reduce the main potential causes, we tend to explain in detail the analysis and realization of the problems using the 5W + 1H methodology, which can be seen in the Table. 3.

From the analysis of the causes of problems in Table 3, furthermore we tend to divide into four (4) classes, then classify the problems and analyze the causes and analyze the possible implications that may occur. After finding the root cause of this problem, the next step is to set a priority scale / RPN (Risk Priority

Number) to improve the implementation of Failure Mode and Effect Analysis (FMEA), supported by many parameters, such as: Severity (S), occurrence (O) and Detection (D), the result are shown as below Table 4.

Table 3: Root cause Analysis

Factor	What	Why	Where	When	Who	How
Man	NA	NA	NA	NA	NA	NA
Material	Auto Deflash	Cutting Edge to Thick	During put the rubber into curing machine	curing process	Operator	Need to modify Mold to get cutting edge 5 micron from outer and inner group
Machine	Auto Deflash	Heater Rod Rusak	perform at curing heating process	curing process	Operator/ supervisor	Need to repair heater gap plate at curing machine
Method	NA	NA	NA	NA	NA	NA

Table 4: Risk Priority Number (from FMEA)

Process	Failure	Effect	potensial root cause	current control	S	O	D	RPN
Curing	abnormal heat transfer (uneven)	Self deflash	Heater Rod Broken	temperature check at beginning of shift	8	4	4	128
	Mold is having a problem	Self deflash	Cutting edge too tick	checking during process	7	6	4	168

From Table 4, we are able to see the failure mode that causes defective/wastefulness and therefore the calculation of the RPN delineate below:

1. Uneven heating within the curing process which ends in auto Deflash / Self deflash

a. Severity 8: High severity, because of the top can feel unhealthy consequences that may not be accepted outside the tolerance limit
b. Occurence 4: Failure is very possible to occur
c. Detection 4: potential causes area moderate unit. The detection methodology is still allow, and so commonly reason for it happens

Based on points a, b, and c: Severity 8, incidence 4 and Detection 4, so the calculation of RPN is a hundred and twenty eight. Because of the multiplication of S, O, and D, as a calculation:

RPN = S x O x D = 8 x 4 x 4 = 128

2. molding issues with the curing process end in AutoDeflash / Self deflash.

a. Severity 7: High severity, because of the top can feel unhealthy consequences that can't be accepted on the far side tolerance
b. occurence 6: The incidence of damage/failure is possible to occur around 0.5%
c. Detection 6: potential causes are moderate, detection strategies still have a potential cause

Based on points a, b, and c: Severity 7, incidence 6 and Detection 4, so the RPN was obtained 168. As it is often from the multiplication of S, O, and D. that is developed as follows:

RPN = S x O x D = 7 x 6 x 4 = 168

Furthermore, as the proposed, it is need to set the priority of the most causes and it was able to put the priority scale of repairing, which has to be done, as follows:

a. Repairing the rubber hardening machine, the Heater Rod.
b. Fixing the Parallelism Plate.
c. Repairs the mold by modifying it, so that the items aren't too thick.

Those 3, higher than the enhancements delineate as shown in Figure 5, 6 , 7 and also table 5 below,

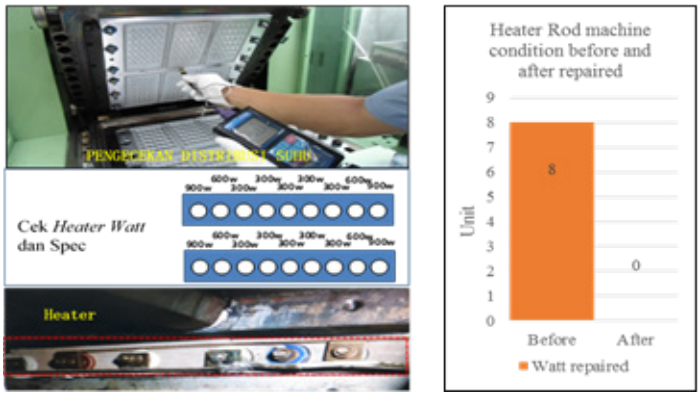


Figure 5: Repairing All Rod Heater Machines

In the heater rod engine repairs are is provided to standardize the wattage heater. That's to induce a normal temperature. A kind of standardization of the Watt Heater. As we can see in the Table 4, above there is a difference in decreasing 8 watts before repairs were made.

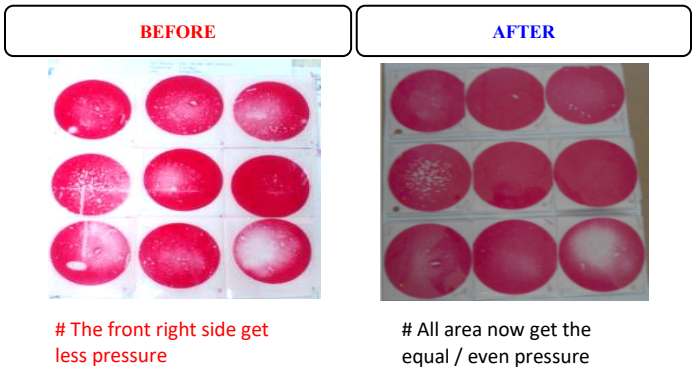
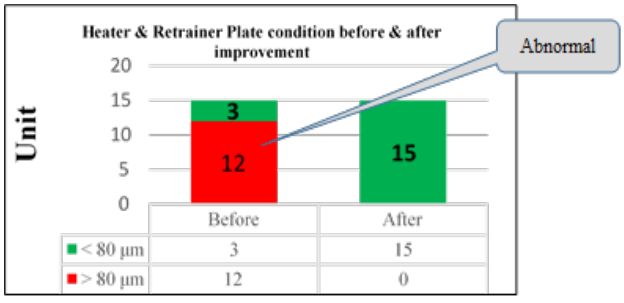


Figure 6: Fixing the Parallelism Plate.

Retrainer Plate Parallelism comparession before and after improvement

Table 5: Heater and retrainer plate condition before & after improvement



Grinding process is carried out on the heater and retrained plate machine to get parallelism, according to the standard (Kumamoto standard parallelism must be <80 um), as shown in Figure 6 and table 5

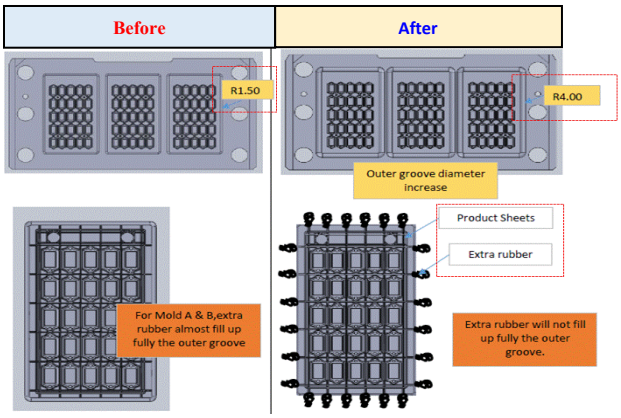
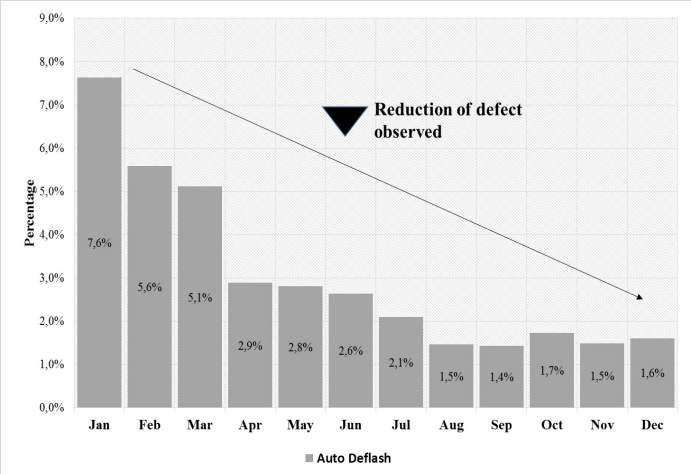


Figure 7: Fix Mold at Outer Groove area

As shown in figure 7. Repaired by enlarging the scale of the outer groove from R1.50 to R4 to manage additional Rubber doesn't fill the outer space of the groove.

Once the improvement on production method was made, research worker conjointly reviews the effectiveness of enhancements from the defective aspect for additional details as we can see on Tables 6;

Table 6: Reducing Trend of defective



Based on Table 6, we are able to see a decreasing in Defective Auto-deflash, up to 80.5% from 11,7% to 1,7% at the end of the year, once repairs were made to the affected processes.

5. Summary and Conclusion

a. The application of FMEA, it will makes the analysis phase more focused, namely by analyzing potential failures, failure effects, frequency of failures and the tools was used to detect failures and the RPN value as priority for improvement as measured by severity, occurrence, and detection so that the priority value obtained is more realistic.

b. From improvements such as defective residual trends and or customer complaints, it shows that improvements made to high quality products have reduced deficiencies. Ishikawa diagrams and FMEA tools are very helpful as a tool for analyzing and characterizing problems and supporting improvements in methods and production.

c. Some technical improvements in certain processes have reduced the level of defects, so these improvements are very effective and useful for increasing company productivity.

Conflict of Interest

I declare that this paper is not having any conflict of interest.

Acknowledgment

Mercubuana University, Indonesia

References

[1] The Register UK daily report. July 2018. 'Western Digital Will Close Factory HDDs in Late Malaysia in 2019.
[2] Kompas.com, Kompas Tekno dari The Wall Street Journal, (2016).
[3] Yamit, Z., Production & Operations Management (First Edition). Yogyakarta: Ekonisia (2011).
[4] Biegel, J. E Pengendalian Produksi-Suatu Pendekatan Kuantitatif, hal 19-63, hal 190-211, Penerbit Akademika Pressindo, Jakarta . (1992).
[5] Purba, H.H., Aristiara, N., and Muda, I., Implementation of Business Process Improvement to Reduce Wastes: A Case Study in Grand Piano Assembly Process. (JSERBR) 2018, 5(5):511-527 (2018).
[6] Ikatrinasari, Z.F, Sawarni Hasibuan, S., and Kosasih, The Implementation Lean and Green Manufacturing through Sustainable Value Stream Mapping. IOP Conference Series Materials Science and Engineering 453(1):012004 (2018).
[7] Hernadewita, Saputra, Y., Yatma, W.A., Prastyo, Y., Irawan, A., Lean Manufacturing Approach for Productivity Improvement using SMED (Single Minute Exchange of Dies), IJRESM Vol 2, Issue 5 May 2019.
[8] Hernadewita and Rohimah, A., Lean Manufacturing Implementation Using VSM to eliminate 7 waste in Painting Process, IJMPRD Vol 9, Issue 2. Apr 2019.
[9] NHSIII, Improvement Leaders Guide. Coventry: NHSIII (2008).
[10] Esmail, A. Patient safety in your practice. Pulse; 71: 3, 22-23 (2011).
[11] Pearson A Minimizing errors in health care: Focusing on the 'root cause' rather than on the individual. International Journal of Nursing Practice;11: 141, (2005).
[12] Scarvada, A.J. A Review of the Causal Mapping Practices and Research Literature. Second World Conference on POM and 15 th Annual POM Conference, Cancun, Mexico (2004).
[13] Tomic, B., and Brkic, V. S. Effective Root cause analysis and corrective action Process, Journal of Engineering Management and Competitiveness (JEMC) vol. 1, No. 1/2, 2011, 16-20, (2011).
[14] Chrysler, C. "Potential Failure Mode and Effect Analysis (FMEA) Reference Manual". Ford Motor Company and General Motors Corporation (1995).

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 361-367 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

The Implementation Model of Character Building in Curriculum 2013 Through The Scout Movement as A Required Extracurricular in Primary And Secondary Education

Lena Nuryanti[1], Deni Darmawan[*,2], Achmad Hufad[3]

[1]*Economic Education Department, Indonesia University of Education, 40154, Indonesia*

[2] *Educational Technology Department, Indonesia University of Education, 40154, Indonesia*

[3] *Community Education Science Department, Indonesia University of Education, 40154, Indonesia*

ARTICLE INFO

Article history:
Received: 04 November, 2019
Accepted: 20 January, 2020
Online: 07 February, 2020

Keywords:
Personality
Character Education
The Scout

ABSTRACT

The character education program conducted through extracurricular areas of the Scout. The method used in this study is the Research and Development (R&D). The research objectives are scouts elementary and junior high school. This research resulted in 3 Prototype Free Model implementation of character education through extracurricular scouting programs in the form of three models that must be followed. All three models include 1) a regular model so-called Block model; 2) Actualization models and 3) Regular models. Having tested through descriptive statistics, the percentage of respondents who obtained the opinion states that the model most widely used is the Block model, which reached 33.78%. The actualization model application reached 6.75%, while the Regular model application reaches 17.57%. The findings further analysis of the model character education through extracurricular activities scouting in basic education in West Java can be done via the Blocks model. The Results of the obtained tendencies that cause suboptimal extracurricular of the Scout to embed the characters early on elementary and secondary school.

1 Introduction

The Preamble of the Constitution of the Republic of Indonesia Year 1945 mandates the Pancasila as the nation and Worldview of Indonesia which should animate all areas of development. One of the very important national development and became the foundation of society, nation, and state is the nation's character building [1].

Paying attention to the circumstances of a national character which tend to be chaos and the poor. The government took the initiative to prioritize the development of national character is supposed to be the mainstream of national development. The national character building means that any development effort should always be considered linkages and their impact on the development of character. The national character building reflected in the mission of national development that positions the character education as a mission the first in line for the purpose of the curriculum in 2013. That is to prepare human Indonesia to have the ability to live as individuals and productive citizens, creative, innovative, and effective, and be able to contribute to the life society, nation, state, and civilization [2].

Minister of Education and Culture explained that the government will make a whole triangle education activity, namely curricular, extracurricular, and co-curricular. "What we create in the curriculum 2013 [3] is a complete triangle, namely knowledge, skills, and attitudes". Build an attitude could not be done only in the classroom, but it formed through extracurricular and co-curricular activities. For this reason, Scout is one of the required activities in extracurricular. Scout movement teaches the value of leadership, togetherness, and has become a phenomenon not only in Indonesia but also in the world. In harmony with the objectives of the Scout is to form each Scout as follows: 1 have a personality who believe, fear Allah, noble, spirited patriotic, law-abiding, disciplined, upholding the noble values of the nation, competent in life, bodily health, and spiritual; 2 be a citizen of the spirit of Pancasila, faithful and obedient to the unitary of Republic of Indonesia as well as a member of the public good and useful, to establish itself independently as well as jointly responsible for the development of the nation and the state, have concern for others

* Deni Darmawan, Email: deni_darmawan@upi.edu

life and the natural environment.

Development of a national character building that pursued by a variety of forms, until now has not done optimally. It reflected in socio-economic disparities; political is still great in corruption, and injustice law, pornography, promiscuity, sexual abuse by teenagers committed to minors, violence, and unrest are likely to become more widespread in many sectors. As a result of the unconformity of stakeholders in all sectors of public life currently prevalent such as anarchy, social conflicts, the narrative of bad language and manners, as well as the disobedience of traffic [4].

Indonesian society is accustomed to polite in behavior, implement consensus in resolving the problem, have the local knowledge-rich plurality, and being tolerant and cooperation is now likely to turn into homogeneous groups to outdo each other and behave dishonestly. All of it was confirmed that the uncertainties identity and character of the nation that leads to 1 disorientation and not understand the values of *Pancasila* as the philosophy and ideology of the nation, 2 the limitations of the policy tools integrated in realizing the values of Pancasila,3 a shift ethical values in the life of the nation, 4 the waning awareness of the cultural values of the nation, 5 the threat of disintegration and 6 the weakening of the nation's independence.

Therefore, in this study want to answer about: "How the implementation of extracurricular education of the scouts as required in primary and secondary education unit listed in curriculum 2013 and the ministry rule No. 81A [3] which further clarified in ministry rule No. 63 of 2014 [5] of Indonesian government on the implementation of compulsory education extracurricular the Scout?"

The desire to do this research came from the discourse will launch Curriculum 2013 [3] at approximately the end of 2012 until approximately the beginning of 2013 which did not escape the turmoil opinion of the pros and cons of teachers, lecturers and education practitioners, especially in the realm of scouting. May 17, 2013, researchers concluded that the results of the meeting to realize that essentially runs the Act No. 12 of 2010 about the scout movement and understand the rules of education and culture minister No. 81A [5]. Therefore, we consider it is necessary to perform research based on the implementation of the curriculum in 2013 [3], especially concerning the compulsory extracurricular of the Scouts.

This research is expected to contribute to national scouts in an effort to undertake the minister of education and culture to implement the 2013th curriculum. This is, of course, the one side is a gift that the Scouts received praise and trust. However, on the other hand, it is a reminder that all components of the Scouts should prepare to deal with it wisely. Considering that the implementation of the exercise scouts does in the "*Gugus Depan*" education units ranging from elementary to groups of "*Siaga*" and "*Penggalang*," "*SMP*" for groups of "*Penggalang*," a high school for "*Penegak*" until College for "*Pandega*." Besides, by the basic principles of scouting embracing the principles of voluntarism, air-shift system, the system of separate units, the system marks the skill and honor code of cropping systems as well as outdoor activities. It is the hallmark, and which is the identity of scouting education.

Therefore, it is important to note that the research team had to observe accurately the formulations and wiser response to the two sides that are considered controversy, i.e. their given "Extracurricular Mandatory" versus "Voluntary Principles" to be a member of the Scouts. The member of the Scouts must be based on the voluntary principles and the word or phrase "love and willingly "this has always been a concomitant sentence pledge or promise dialogue participants before the coach when sworn or affirmed as a member of a variety of requirements.

Outcomes of the research are expected to show a prototype either manual or reference shape model of character education through Scouting as an extracurricular compulsory education that has been tested. This is expected to provide at least a minimum contribution of reference for national education units and education units as a forwarding scout group basis. In the channel used is the path began the *National Kwartir*, *Regional Kwartir*, *Branches Kwartir*, *Twigs Kwartir*, and *Gugus Depan* because the next group is spearheading the efforts to establish the character of their students to schools throughout Indonesia.

The position of the *regional kwartir* of the Scouts movement in West Java in the activities of this study is used as a group control of information. In connection with the Institution of *Universitas Pendidikan Indonesia* (UPI), wherein Chief Researcher and study members were lecturers of UPI included Cluster front of existing and based in UPI be under the coordination of *regional kwartir* of west java [6].

2 Curriculum Development in 2013

Today the growing demands for changes in the education curriculum emphasize the need to build national character. It based on the facts and perceptions about the declining quality of the attitudes and morals of children or young people. Curriculum 2013 [3] is prepared to print a generation to be ready in facing the future because the curriculum is structured to anticipate future developments. The development of a national character must be actualized substantially in the form of national action to establish the spiritual, moral, and ethical.

This performed through a systematic and integrative approach with the involvement of the family, education units, government, people including peers, young people, the elderly, the mass media, the Scouts, community organizations, political organizations, professional organizations, non-governmental organizations, strategic groups such as structural elite, the political elite, journalists, cultural, religious leaders, traditional leaders, and community leaders.

As for character development strategy, it can be done through awareness, education, empowerment, acculturation, and cooperation with due regard to environmental conditions and community needs as well as a multidisciplinary approach that does not emphasize indoctrination. As mandated by Law 20 of 2003 [7] on National Education Systems in the explanation of Article 35, where the competence of graduates is qualified to graduate capabilities that include attitude, knowledge, and skills by the national standards that have been agreed. Moreover, in the curriculum included, the Scout is compulsory in every school. Scout through the Homeland will be maintained intact.

Curriculum development that occurred in Indonesia is already happened several times, wherewith the curriculum in 2013 [3] is already five times. For more details of how the development of the curriculum of every turn, the shape change of the curriculum that occurred in Indonesia specified in the curriculum of public test material 2013. Triangle Whole Curriculum 2013 [3] to a

target the new philosophy that proclaimed Indonesian education as an integral part of the learning process of children. The education ministry of Indonesia said that the form of the whole triangle is curricular, co-curricular and extra-curricular. Curricular activity is an activity that must be followed by learners and binding. The curricular program contains a wide range of fundamental skills and abilities students need to have a minimum at level schools (educational institutions).

Co-curricular activities are an integral part of teaching and learning (curriculum) which serves as a support and enrichment curriculum for life skills coaching implemented gradually and continuously. The extracurricular program is an internal part of the learning process that emphasizes meeting the needs of learners. Between curricular and extra-curricular activities cannot be separated, even extra-curricular activities complementary extension or reinforcement curricular activities to channel the talent or driving the development of potential learners to achieve the maximum level. The Scout movement is an integral part of the educational process to prepare to practice a code of honor Scout: "*Satyaku ku dharmakan*" (my promise is my behave), "*Dharma ku kubaktikan*" (I actualize my behave). The code is to instill the confidence, adding to the spirit of service to the nation and state, the practice of the Scout's *Tri Satya and Dasa Dharma* has a work culture that based on devotion to the noble and sincere. Thus, the intact triangle of the curriculum in 2013 [3] will be homework for teachers who spearhead education.

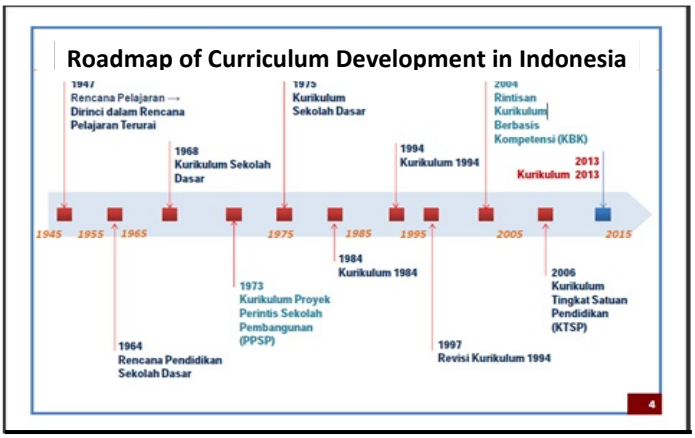


Figure 1: The development of the curriculum in Indonesia [8]

This is presumably as a counterweight formal learning activity in the curriculum which more oriented to the cognitive (knowledge) and psychomotor (skills). The Scouts will be able to build the intelligence of the students in the realm of affection (attitudes and behaviors) so that students will be able to develop a positive character.

3 Meaning of the Character Building

Character means behavior or personality. Hill T.A [9] says that "Character determines someone's private thoughts and someone's action done. Good character is the inward motivation to do what is right, according to the highest standard of behavior in every situation".

In this context, the characters can be interpreted as someone's identity. According to Ekowarni [10] on the order of micro, the character is defined as (a) the quality and quantity of the reaction against oneself, another person, or a particular situation; or (b) the character, character, psychological traits. Psychological characteristics of the individual in the private sphere, in evolution will develop into characteristic groups and, more broadly characterize the social. Individual psychological characteristics will give color and patterns of group identity and in the framework of the macro will be a psychological trait or character of a nation.

Based on the above understanding can be said that the character is an identity, personality, and character are inherent in a person. Characters are always related to the physical and psychological dimensions of the individual. The character is the value of the basic behavior of the reference values human interaction (*when the character lost then everything lost*).

4 Definition of Education Extracurricular

About Permendikbud No.81a later confirmed in more specific in Permendikbud No. 63 of 2014, about the implementation of Extracurricular Mandatory Scouting, it is important to know what the sense of Education Extracurricular as follows [3]. Education Extracurricular is "educational activities undertaken by students out of hours learning the standard curriculum as an extension of curriculum activities and carried out under the guidance of the school with the aim of developing personality, talents, interests, and abilities of learners wider or outside interests developed by the curriculum" [11].

The function of the extracurricular the Scout activities includes (1) The function of the development, which supports personal development through the expansion of interests, development potential, and providing opportunities for character formation and leadership training. (2) The social function, namely to develop the ability and internalization of moral and social values, (3) Function recreational, which is carried out in an atmosphere of relaxed, happy, and fun, (4) Function career preparation, namely to develop career readiness of learners through capacity building.

5 Research Method

This study uses a quantitative research and development approach (Research and Development). In the opinion of Gall et.al (1996) [12], the research has three main steps, namely: a preliminary study, stage of development, and testing phase. In education, the R&D process has been used in educational research related to educational practice and has published some research findings into educational products that are applicable for the purpose of improvement of learning in the school environment. Process R & D has been used to test the predictions and to develop effective instructional interventions that have been used to improve practice in the classroom. As suggested by Gall et.al (1996) [12], the original ten-step R & D process that has been embraced by the industry, while for research in the field of education has been limited to seven first steps.

Researchers, in this case, will be followed closely in the years to one with a seven-step process of testing, evaluation, and improvement as it has produced valid product research (See Figure 2). R&D Process "…*consists of a cycle in which a version of the product that developed, field-tested, and revised by field test data*". explained that the R&D consists of a cycle of "product was

developed, field-tested, and revised based on the test data field and has produced educational products are fully ready for operational use in schools."

Step one to seven R&D processes are includes: (1) the analysis of the research and proof of concept, (2) the product planning and design, (3) product development, (4) field test early, (5) the revision of the prototype, (6) major field test, and (7) the revision of the final product (Borg & Gall, 1996 pp. 784-785). [12] Figure 1 graphically described the process for this study.

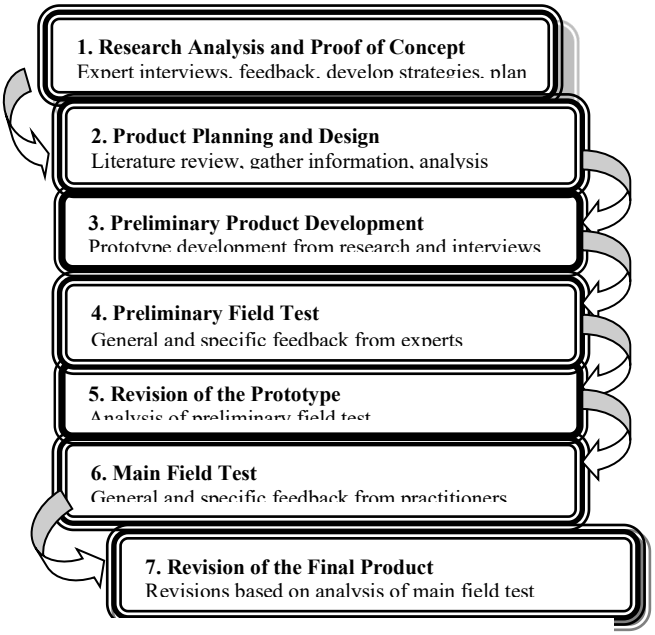


Figure 2: Research Product (Borg & Gall, 1996 hal. 784-785) [12]

6 Data Collection Techniques

The collection of data that researchers use to collect information and data required by the technique as follows:

1) Study Library, which is a technique of data collection is done by collecting, reading, and studying a variety of literature and reading materials related to the problem being studied.
2) Field Studies, namely the collection of data obtained directly, through the following activities:
   a. Interviews, namely data gathering activities were carried out to gather information and data with a list of questions prepared in advance in which the questions are asked of informants.
   b. Focus Group Discussion (FGD), the activities carried out in the form of a meeting organized by the researchers by inviting competent parties (stakeholders) or the *Regional Kwartir* and ranks mainstay and coaches scoutmaster required in the study.
3) The collection of additional data and the formulation of the early stages of design through activities Seminar and Workshop.

7 Processing and Data Analysis

To undertake the processing and analysis of data obtained from qualitative research, Sugiyono [13] states:

> *Data analysis is the process of systematically searching and arranging the interview transcript, field notes, and other materials that you accumulate to increase your understanding of them and to enable you to present what you have discovered to others.*

The opinion makes clear that the activities of data analysis are a systematic process of searching for and compiling the results of interviews, notes, and other resource materials are combined so easily understood and what is gained can be presented to others. Thus, it can be understood that the processing and data analysis is the process of searching and compiling the data obtained as a result of observations, such as notes or interviews. They are arranged systematically by choosing what's important and what is to be learned so that they can be made a conclusion that can be understood by researchers themselves or others. In this study, data will be collected from some respondents. A grating instrument is created to collect data on variations of the model design prototype design for the implementation guide of the extracurricular scouting.

8 Results and Discussion

Here are the findings regarding the implementation of the Scout as an extracurricular education compulsory at Primary and Secondary Education unit listed in Curriculum 2013 and in Ministry of Education Rule No. 81A which is further clarified in Ministry of Education Rule No. 63 of 2014 on the implementation of Compulsory Education Extracurricular of the Scout.[3] At the level of schooling, the implementation began in the following stages:

8.1. Analysis of the implementation level extracurricular of the Scout

Based on the findings in the field most of the effort to analyze the importance of the Scout in the cluster (*gugus depan*) level is quite high. These findings are characterized by the number of respondents who stated that The Scout is compulsory in school extracurricular. Various preparations have been done in order to expedite the extracurricular activities, particularly the development of the capacity of a manager or instructor also suggestions infrastructure to support the scouting activities. The Scout extracurricular programs can be followed by all students, either normal students (physically) or those who have disabilities or mental disorders. So, all students have the opportunity to follow the Scout extracurricular program, where all students are equal and there are no differences. In findings regarding equal rights for students to join the program supports the idea of the results of research that has been done by Hong-Joong Cho, Jung-Ran Kim, and Sang-Gu Kim (2015) [14] which states that “Considering our educational climate that difference easily return to superiority or inferiority in our society, two cases of freedom of speech and discriminatory speech in Japan have great implications for us.

8.2. Legal formal school level in implementing the compulsory education of the Scout extracurricular.

Based on the research results show that the organizing extracurricular of the Scout at the school level requires decree on mandatory legal standing as for whether or not they are carried.

The research findings indicate that most respondents from the school state that the decree is important. A decree or legal standing has been owned by the schools in Indonesia, thereby the Scout extracurricular activities are mandatory for all students. To meet the liability aspect of this that the schools should be able to perform as well as possible. This model is a model that is the most beneficial aspect of effectiveness in applying the values of independence, chivalry, honesty, and cultivation of the soul "progressive persist". From this model, the implementation of infrastructure support has been prepared by the schools with adequate facilities and infrastructure. During the implementation of scouting activities then the condition of moral, spiritual must be balanced with the support of the material and financial aspects were adequate. During the execution of an Extracurricular scouting program, the direct leadership of the principal as to be able to control it well, given good leadership at the school level and even universities will be required to oversee all activities until successful. This finding is consistent with the results of previous research conducted by M. Alsharari Mufadhi Ratyan and A. Almadani Maen Moh (2015) [15] which states that "Deans are controlling a work for teaching and scientific research in colleges, leaving all faculty members them with less time in serving the community. And enhancing social responsibilities approaches.

8.3. Development of Systems and Structures Extracurricular of the Scout

After the implementation goes well, then the coaching process performed by the scoutmaster at each school. Most of the coaching process has become an important note. Support for extracurricular coaches is done in the form of facility fees, awards, duties, and responsibilities, as well as the cumulative value of the ranks in his career as the coach. To supervise the coaches are often accompanied by parents during the execution of certain activities. Meanwhile, to obtain legal professionals in the field of scouting is then formed *kwartir* branches at the district level, this is done through the process of coaching *kwartir* Trustees candidate school level and also formed a team of policy dissemination extracurricular scouting. While at the district level branches formed *kwartir* that work to training, socialization process of the formation of formators for *kwartir* and school level.

Basically, the development of learning systems can support the planting of noble values taught in the Scouting extracurricular activities, such as can be done with the e-learning system. Through e-learning is the process of coaching can be done anywhere and anytime, it supports the findings of research carried out by V. Murugananthan and B. L. ShivaKumar, (2015) [16] e-learning methodology includes Multimedia learning, technology enhancement, computer-based instruction, computer-based training, computer-aided instruction, web-based learning, and virtual learning. Through these kinds of systems that learning the scouting extracurricular activities could be more widespread and freely followed by all students anywhere and anytime.

8.4. The Model of the Implementation and Evaluation to the Extracurricular at School

The implementation process extracurricular education compulsory in every school scouting majority combined model block, actualization, and regular). Given these three models is quite helpful in monitoring the implementation of compulsory education extracurricular scouting done directly or indirectly. To evaluate the process of implementation of compulsory education extracurricular scouting implemented through reporting by the builder school level. All three models support the implementation of all the policies formulated in the General Assembly's decision of the Scout Movement of Indonesia (*Keputusan Munas Gerakan Pramuka Indonesia*) [17].

During the implementation of extracurricular of the scout programs in schools, they are supporting and which have an impact on the success of this program. Figure. 3 is an overview of the number of supporters and obstacles experienced by schools in Indonesia.

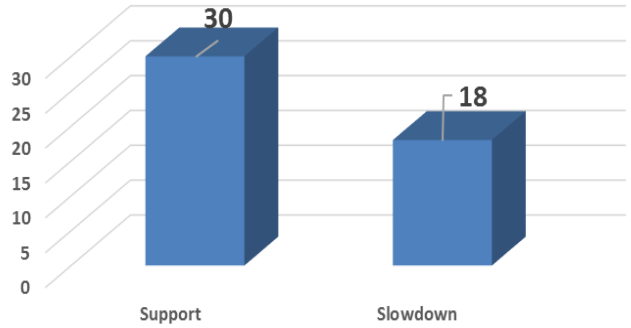


Figure 3: The Scout Implementation at School

8.5. Implementation of Government Regulation No. 63 Year 2014 [18] concerning Extracurricular of the Scout in Schools.

The implementation of Law No. 63 of 2014 [18] on the implementation of scouting in school extracurricular otherwise required by approximately 39 schools from 40 school respondents, while 34 schools agreed to continue to apply. In addition, 38 schools out of 40 schools expressed quite relevant and there are 30 schools expressed sufficient support to the success of this program. Figure.3 is an overview of the results of this study.

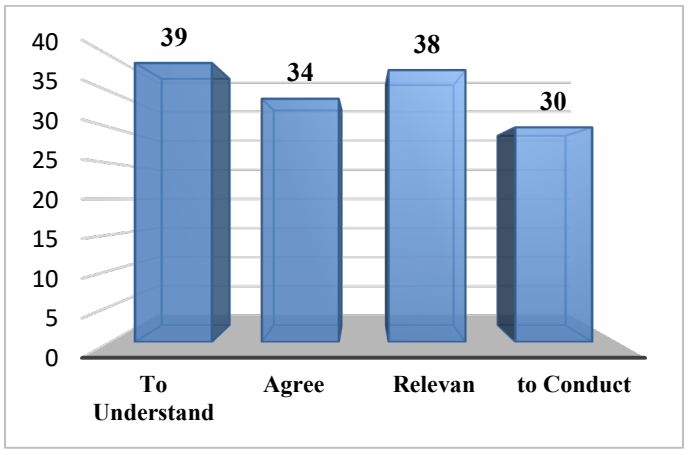


Figure 4: Overview of Implementation the Scout Program based on Government Regulation No. 63 Year 2014 [18]

Successful implementation of the Scout extracurricular program has been basically set in *Kwartir Nasional* (*National Quarter*) decision[13], where each Scout institution organizes the program is obliged to provide support facilities, infrastructure, human resources, as well as adequate financing. The findings of this

study indicate that the aspect of comprehension, expression agree, relevance and support school institutions is very good. Actual implementation instructions scouting program has been set out clearly by the national quarter of the Scout movement [19].

9 Measurement of Opinion Successful Implementation of Law No. 64 Year 2014 [18] concerning Extracurricular School.

Here is the condition of the successful implementation of Law No. 63 Year 2014 [15] regarding the implementation of extracurricular scouting at the school level. Where the results showed that as many as 29 states are needed for implementation in schools. As for most of that during this Law 63 successfully executed well by the coaches at the school, about 35 people express this condition.

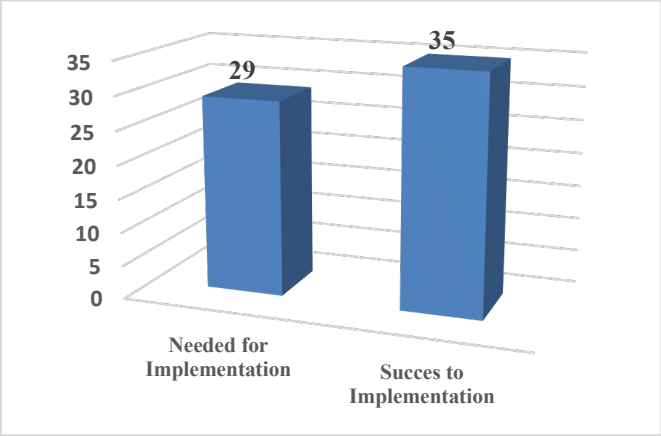


Figure 5: Data Measurement of Implementation The Government Role No.63 Year 2013 [17].

The success of the school in the Scout extracurricular programs, at least be able to promote the quality of service and guidance that has been done to other schools. This is, of course, will require education program marketing strategies appropriately. As a reference to the model of digital marketing through the website of the school in the education of which can be adopted on the findings of research conducted by Tripti Dhote, Yatin Jog, Nutan Gavade and Gesu Shrivastava (2015) [20] which states that "any form of direct or indirect marketing that is used to build awareness, recognition, recall, and action for a brand, business, product, person, or other entity and is carried out using the tools of the social Web, such as blogging, micro-blogging, social networking, social bookmarking, and content sharing". From those opinions, it can be recommended that the number of results and the strength of the Scout program should be packaged and disseminated through the online system of the school website. Thus, the success of the inter-school will rapidly grow together in an effort to instill the values of national identity through the Scout extracurricular program.

10 Conclusions and Recommendations

10.1. Conclusion

Implementation of the extracurricular field scouting program in the school can be done by providing socialization of the importance of such activities. Then set up a training program for prospective Pembina county level, district level and school level. Having formed the structure of Trustees of the management structure of the system is built coaches for the district, district and school. Once the structure and team coach this program, there was a dissemination of registration and membership training scouting. Every step is formulated in the extracurricular scouting program for this in Indonesia. There are a number of conditions that need to be considered in the implementation of this program, in particular, support for the program, Pembina power, infrastructure, financing. Meanwhile, to measure the success of this program required a clear law of the government, in Indonesia owned by Law No. 63 of 2014 [14] as an umbrella implementation scouting extracurricular activities at school. During this time the law provides enough ease and success as well as the convenience of the implementation of the extracurricular program.

10.2. Recommendation

The Scout activities can be maintained as an extracurricular program that is capable of forming the character of the nation as a knowledgeable society and independent. In an effort to maintain a program that has been successful, thus the government regulation No. 63 [14] of the Implementation Program Extracurricular must be maintained by the school, district, and county. This needs to be supported by the government, schools, people and concerned citizens in efforts to form a national character

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors gratefully acknowledge that this work has been supported by a grant from the Ministry of Education and Culture of Republic Indonesia.

References

[1] Muhamad, B., & Saparahayuningsih, S, "An Attitude and Character Instructional Development Based on Curriculum 2013 in Elementary School" Creative Education, 07(02), 269–277, 2016. https://doi.org/10.4236/ce.2016.72025
[2] Njikeh, K. D, "The Mentally Offset Human and the Society" Open Journal of Psychiatry, 06(02), 188–190, 2016. https://doi.org/10.4236/ojpsych.2016.62022
[3] Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 81A of 2013 concerning the Implementation of Curriculum, 2013, pp. 1-97.
[4] Zhang, J. "Analysis on Standardization and Individualization of Nation Wear" Art and Design Review, *06*(02), 96–107, 2018. https://doi.org/10.4236/adr.2018.62009
[5] Law of the Republic of Indonesia Number 12 of 2014 concerning the Scout Movement, 2010, pp. 1-23.
[6] Scout Research and Development Center, West Java Regional Kwartir, Research Results, 2011, pp.22.
[7] Law of the Republic of Indonesia Number 20 of 2003 concerning the National Education System, 2003, pp. 1-26.
[8] Law of the Republic of Indonesia Number 12 of 2010 concerning the Scout Movement, 2010, pp. 1-23
[9] Hill, T.A., 2005. Character First! Kimray Inc., http://www.charactercities.org/ downloads / publications / Whatischaracter.pdf. downloaded on August 11, 2011.
[10] Ekowarni, Endang. Development of Budi Pekerti's Noble Values as National Character, 2009.
[11] Suyatno, Scouting Education through Education Unit-Based Clusters, East Java Pusdiklatda, Surabaya, 2014.

[12] Gall M, Borg W, Gall J. Educational research: An introduction. White Plains, NY: Longman, 1996.
[13] Sugiyono, Educational Research Methodology, 2007.
[14] Hong-Joong Cho, Jung-Ran Kim, Sang-Gu Kim, "Special Education Discourses on Freedom of Speech and Discriminatory Speech", Indian Journal of Science and Technology, 2015 Mar, 8 (S5), Doi no: 10.17485 / just / 2015 / v8iS5 / 80027.
[15] M. Alsharari Mufadhi Ratyan, A. Almadani Maen Mohammad, "Saudi Higher Education Reality and Prospects: Evaluating Careers' Dimensions of University Teaching, Scientific Research and Community Service Northern Border University as a Model", Indian Journal of Science and Technology, 2016 Jan, 9 (4), Doi no: 10.17485 / ijst / 2016 / v9i4 / 84984
[16] V. Murugananthan, BL Shiva Kumar, "An Adaptive Educational Data Mining Technique for Mining Educational Data Models in Elearning Systems", Indian Journal of Science and Technology, 2016 Jan, 9 (3), Doi no: 10.17485 / ijst / 2016 / v9i3 / 86392
[17] Decree of the National Conference on Scouting Movement Number 11 / Munas / 2013 concerning the Statutes and By-Laws of the Scouting Movement, 2013.
[18] Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 63 the Year 2014 Regarding Scouting Education as a Compulsory Extra-Curricular Activity in Basic Education and Secondary Education, 2014, pp.1-4
[19] Decree of the National Kwartir Number 231 of 2007 concerning Guidelines for the Implementation of the Front Scout Movement, 2007.
[20] Tripti Dhote, Yatin Jog, Nutan Gavade, Gesu Shrivastava, "Effectiveness of Digital Marketing in Education: An Insight into Consumer Perceptions", Indian Journal of Science and Technology, 2015 Feb, 8 (S4), doi no: 10.17485 / ijst / 2015 / v8iS4 / 70300.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 368-372 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

The Proportionality of Women Graduated from the Professional Career of Mechanical and Electrical Engineering at UNTELS: Analysis of their Academic Performance and Labor Field of Action

Dora Yvonne Arce Santillan[1,*], Omar Freddy Chamorro Atalaya[1], Yesica Pamela Leandro Chacón[1], Jorge Isaac Castro Bedriñana[2], Elizabeth Rosario Martínez Santillán[3]

[1]*Faculty of Engineering and Management, Universidad Nacional Tecnológica de Lima Sur Lima, Lima, Perú*

[2]*Faculty of Zootechnics, Universidad Nacional del Centro del Perú, Huancayo, Perú*

[3]*Faculty of Medical Technology, Universidad Nacional Federico Villareal, Lima, Perú*

ARTICLE INFO

Article history:
Received: 20 November, 2019
Accepted: 28 January, 2020
Online: 07 February, 2020

Keywords:
Academic performance
Graduated Women
Mechanical and Electrical Engineering
Development Field
Mechanics Subjects
Electrical subjects

ABSTRACT

This research seeks to describe the proportionality of women graduated from the Professional Career of Mechanical and Electrical Engineering, analyzing their academic performance and the field of work action in which they are working. This study was done with the purpose of highlighting academic performance, and encouraging women to continue with this career, eliminating the myth that Mechanical and Electrical Engineering is not for women; In addition, this study will be a source of information for the National Technological University of Lima Sur, since, with this, action plans can be generated to increase the number of women entering in the coming years. Once the research has been carried out, it is concluded that the percentage of women who graduate in relation to the percentage of men, per academic semester 2012-II to 2018-II, is equal to 8.1%. Likewise, the weighted average of the academic performance of women graduated in the professional career of Mechanical and Electrical Engineering, is 12.98, this represents a high average for the career; so it can be indicated that the 38 women graduated, have had a good academic performance. Finally, it is concluded that there is a high Spearman ratio of 0.822 between the weighted average by specialty with the field of labor action; These results agree with what has been indicated in relation to the fact that there is a higher qualification in the specialty of Mechanics (AM) with a weighted average range of subjects from 11.73 to 15.42, and according to professional practice reports, of the 38 women graduated from the Mechanical and Electrical Engineering career, 25 are working in the field of Mechanical specialty. With these results it can be said that although the percentage of women linked to this career is still low; Women in Mechanical and Electrical Engineering have a good academic development, which is why women should continue to be encouraged to be included in the area of science, technology and innovation.

1. Introduction

Women represent 52 percent of the world's population, but do not appear to be included in the careers of science, technology, engineering and mathematics, remaining as a mostly male reducer; starring less than 20% of university enrollments, however the historical trend has been reversing and increasing female participation, especially in some specialties of engineering.

In recent years, there has been an increase in the professional development of women in engineering careers. This is because little by little, myths and beliefs about the position that women have in society have been left behind. [1] Although this insertion in this sector has been difficult for women, they have been placed in different areas where their potential can be exploited.

Women participate as men in the fields of training and especially in the university, but the distribution by career of both sexes is very different. Women opt for careers related to the humanities, education or health sciences to a greater extent. The careers that are within the range of "hard", such as engineering, remain poorly chosen by women, remaining as a mostly male stronghold. [2]

* Dora Yvonne Arce Santillan; Sector 3 Grupo 1A 03, Av Central, Villa El Salvador - UNTELS. doratella_92@hotmail.com

Thus, far from being able to say that the low percentage of women who decide on engineering careers is something the result of the late incorporation of women into the university, and the world of work in general, and that it will increase as normalized The situation of equality in society, the number of women who opt for engineering continues to decline, regardless of whether women are achieving effective equality in all other aspects of society. [3]

However, it is also recognized that female participation in higher education in Latin America follows in general terms, a pattern that favors some areas of knowledge over others. [4] However, it should be noted that while female participation in areas such as industrial engineering and computer engineering increases, other areas such as mechanical engineering and electrical engineering are still considered as low or nil preference of women.

It is probably because of this, that the work of women is strongly related to the secondary, "supportive" of men, undervalued and undervalued. This means for women a double challenge, on the one hand, joining the work environment and, on the other, obtaining social recognition for their professional performance. [5]

Within these implicit barriers, there are "masculine" standards associated with the scientific career that women must achieve, whether these, for cultural historical reasons, or also associated with the social division of labor. This is how for women scientists the possibility of reaching these standards, is usually, once they have finished with the vital stage of reproduction and / or upbringing, and there arises the problem that is generated by the difference in the years of trajectory. [6]

Peru is no stranger to this problem. According to the National Superintendence of University Education (SUNEDU), it indicates that, in Peru, the total undergraduate graduates in 2016 were 110,408, of which 54.3% were women. However, the largest number of graduates are in the careers of administrative and commercial sciences and in economic and accounting sciences. Of the total graduates, only 26% were from careers linked to the CTI. Of the total CTI graduates, only 32% were women. [7]

Women are increasingly expanding their field of development related to science, technology and innovation. [8] Engineering is no stranger to them, although, the percentage of women linked to the Mechanical and Electrical Engineering Degree is still low.

A study by The National Academies of the United States has found that social stereotypes condition the academic performance of women, leading to a drop of up to 90% in studies when they are not adequately valued. [9] These prejudices cause math and engineering careers to be generally discarded by high school students, which means a loss of talent for society.

2. Methodology

2.1. Kind of investigation

This research is descriptive. [10] Initially it is intended to identify the percentage of women who graduate in relation to the percentage of men, by academic semester, it should be noted that the National Technological University of South Lima has a total of 15 promotions; In addition, the academic performance of women in the professional career of Mechanical and Electrical Engineering will be analyzed, and it will be determined if there is any type of association between the weighted average by specialty (Mechanical Subjects or Electrical Subjects) with the field of labor action where Currently, women graduated from the degree are working.

2.2. Population and Sample

The population of this research is represented by the 38 women graduated from the School of Mechanical and Electrical Engineering of the academic semesters 2012-II to 2018-II; It should be noted that the first promotion of graduate students of the university occurred in the academic semester 2011-II, but to date there were no women in the career. It also indicates that women represent 8.1% of the total graduates.

In that sense, the sample will be equal to the population [11]; due to the fact that the sample is less than 50, that is, the general weighted average and the weighted average by specialty (Mechanical Subjects and Electrical Subjects) of the 38 women graduated from the School of Mechanical and Electrical Engineering, until the 2018 II. Below is shown in table 1, the proportion of women and men graduated from the career of Mechanical and Electrical Engineering from 2011-II to 2018-II.

Table 1: Proportion of women and men graduated

Graduated Men	429	91.9%
Graduated Women	38	8.1%
Total	467	

UNTELS

2.3. Measuring Instruments

The instrument used is the technical reports [12], of the notes of the graduates of the professional school of Mechanical and Electrical Engineering; academic records, contained in the university management computer system (SIGU) of the National Technological University of South Lima, were used as a source; in the case of the field of labor development, the files are required as a source, which work in the university's degree and degree office, which requests the graduates as one of their requirements to submit reports of professional practices, in which Specify the company and the field of development.

3. Results

Figure 1 shows the number of women and men who graduate per academic semester, from the semester 2011-II to 2018-II.

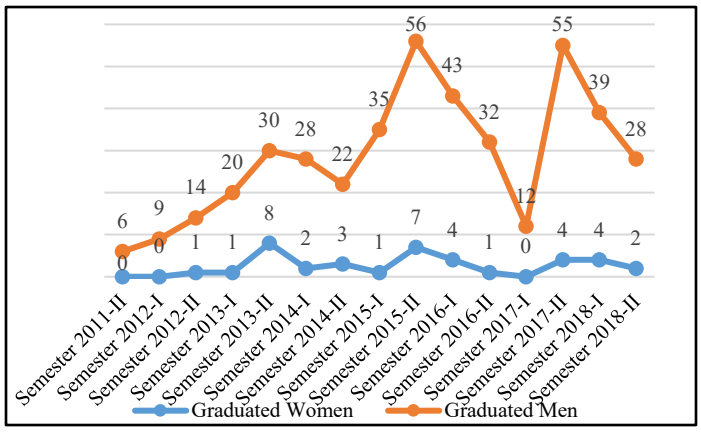


Figure 1: Proportion of women and men graduated by academic semester

As shown in the previous figure, there is a greater number of men graduated in the professional career of Mechanical and Electrical Engineering, as shown in Table 1, this proportion is 8.1% of women graduated and 91.9% of men graduates

Continuing with the investigation, in the following Table 2, the subjects belonging to the areas of specialty are shown; both Electrical and Mechanical subjects.

Table 2: Subjects by specialty areas

Asignaturas del Área de Mecánica		Asignaturas del Área de Eléctrica	
AM1	Industrial Manufacturing Processes	AE1	Electrical circuits
AM2	Industrial Manufacturing Process Laboratory	AE2	Electrical Circuit Laboratory
AM3	Applied Turbomachines	AE3	Electric machines
AM4	Heat Transfer	AE4	Laboratory of Electrical Machines
AM5	Welding Technology I (Corrosion)	AE5	Electrical Measurements
AM6	Design of machine elements	AE6	Electrical Measurement Laboratory
AM7	Internal combustion engines	AE7	Electrical installations
AM8	Industrial Machines and Systems	AE8	Control and Control Board Design
AM9	Maintenance Engineering	AE9	Energy Transformation
AM10	Mechanical Engineering Laboratory	AE10	Renewable Energies I
		AE11	Power systems
		AE12	Thermoelectric plants
		AE13	Lines of transmission
		AE14	Electrical Protection Systems
		AE15	Hydroelectric Power Stations

UNTELS

In order to determine if there is any type of association between the weighted average by specialty (Mechanical Subjects or Electrical Subjects) with the field of labor action where the 38 women graduated from the career currently work, then in Figure 2, the weighted average of the mechanics subjects is shown, while in Figure 3 the weighted average of the electrical subjects is shown.

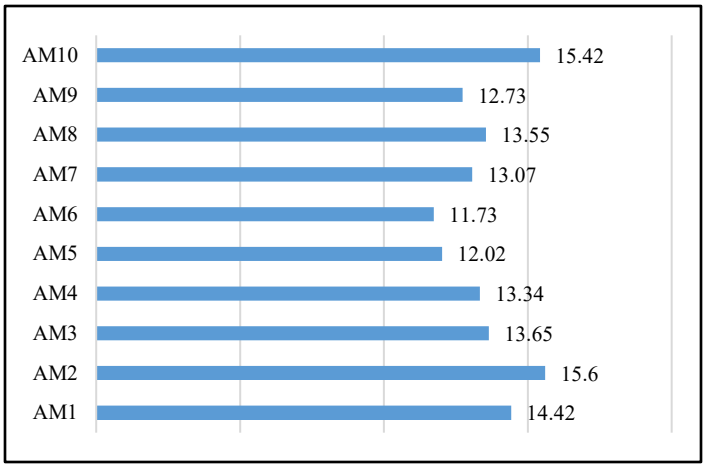


Figure 2: Weighted average of the subjects of the specialty of mechanics

As it is observed, in the previous figure the average range of the subjects of the specialty of mechanics (AM) is from 11.73 to 15.42; obtaining the highest grade the mechanical engineering laboratory subject; being this course of a practical nature, it provides the student with tests in energy transfer and transformation facilities; Likewise, the interpretation of the results of the experimental tests of the mechanical quantities is disclosed, through fundamental instruments and equipment in engineering.

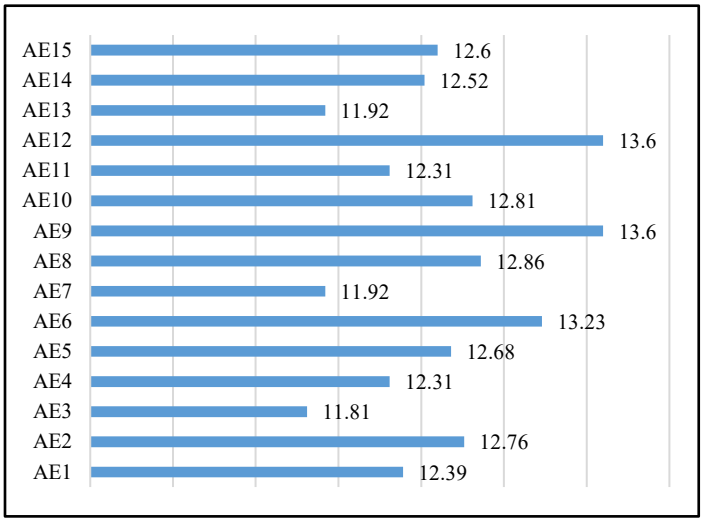


Figure 3: Weighted average of the electrical specialty subjects

As can be seen, in the previous figure, the average range of subjects in the specialty of electrical (AE) obtained by women graduates is 11.81 to 13.6; obtaining the highest average subjects of Thermoelectric Power Plants and Energy Transformation; being this course of a practical nature; The main objective of the course of thermoelectric plants is to show the state of the art in combustion technology for electricity generation, and the basic knowledge of economics and environment. While in the course of energy transformation it is sought to show the definitions related to the ability to work, transform or set in motion by means of the energy involved, for example in rifle fuels.

With the results obtained, it can be determined that there is a higher qualification in the specialty of Mechanics; Likewise, the general weighted average of both specialties was obtained, of the 38 women graduated, being 12.98; said average is considered high for this professional career; so it can be indicated that the 38 women graduated from the Mechanical and Electrical Engineering career have had a good academic performance.

4. Statistical Analysis

Next, the relationship between the weighted average by specialty (Mechanical Subjects or Electrical Subjects) with the labor field of action of the 38 graduates of the Mechanical and Electrical Engineering career will be determined; by means of the Spearman correlation test, which was carried out with the statistical program SPSS version 25. It should be noted that this test was used because the data does not have a normal distribution. [13] The following table shows the test results.

As the Spearman correlation coefficient is shown in the previous table, it is equal to -0.822. This result can be interpreted in this way: There is a high indirect and significant relationship

between the weighted average by specialty with the field of labor action. [14]

Table 3: Spearman correlation

Rho de Spearman			
		Average Specialty	Labor field
Average Specialty	Correlation coefficient	1,000	-,822**
	Sig. (bilateral)	.	,000
	N	38	38
Labor field	Correlation coefficient	-,822**	1,000
	Sig. (bilateral)	,000	.
	N	38	38

**. The correlation is significant at the 0.01 level (bilateral).

SPSS

These results agree with what has been pointed out in relation to the fact that there is a higher qualification in the specialty of Mechanics and according to the reports of professional practices, of the 38 women graduated from the Mechanical and Electrical Engineering career, 25 are working in the field of Mechanics specialty. The following figure shows the percentage representation of the proportion of working in the labor field of the 38 graduates.

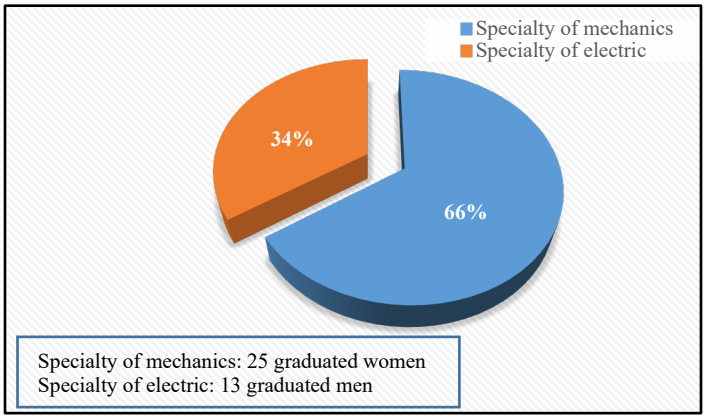


Figure 4: Percentage of the development in the field of labor action

These results agree with what has been pointed out in relation to the fact that there is a higher qualification in the specialty of Mechanics and according to the reports of professional practices, of the 38 women graduated from the Mechanical and Electrical Engineering career, 25 are working in the field of Mechanics specialty. The following figure shows the percentage representation of the proportion of working in the labor field of the 38 graduates.

The reliability criterion of the instrument was determined by Cronbach's alpha coefficient, developed by J. L. Cronbach and recommended for Likert-type attitude scales; [15] through the reliability test with the statistical program SPSS version 25, the reliability of the instrument will be determined.

Table 4: Reliability test

Cronbach Alfa	N of elements
,930	2

SPSS

As can be seen, Cronbach's Alpha is equal to 0.930, so it can be affirmed that there is a "Very High Reliability and Validity of the instrument", since this value is within the range of 0.90 to 1.00. [15]

5. Discussion

In relation to the results of this research, the following discussions are held below:

Regarding the percentage of women who graduate in relation to the percentage of men, by academic semester of the Mechanical and Electrical Engineering degree of the National Technological University of South Lima, which is equal to 8.1%; This result is similar to the research carried out at the Universidad de los Andes, points out that the world percentage of women graduated from the engineering area represents 7%; This is because currently still the inequality and stereotypes that still prevails in these areas, this discourages women in the decision to pursue engineering careers. [16] Likewise, in Peru, only 25% of students in Engineering careers are women, according to SUNEDU, and of this percentage, less than 5% opt for specialties called "hard", such as Mechanical Engineering, Electronics, Civil or Mining. The gender gap in science, technology and innovation careers, which is a worldwide trend, is the result of discrimination against women in these areas of study. [17]

In relation to the academic performance of women in the professional career of Mechanical and Electrical Engineering, which is 12.98, this represents a high average for the career; This result is similar to the research carried out in Spain, which indicates that 13.5% of women obtain the highest qualification compared to 10.1% of men [18]; In spite of this, there is still inequality in remuneration, and in companies there is a higher rate of hiring of male engineers. There is another fact that reaffirms the best performance of women. In the UTN of Buenos Aires when the best averages are analyzed, 50% are of the girls, this because they have more dedication, perseverance and responsibility. [19]

Regarding the high relationship that exists between the weighted average by specialty (Mechanics Subjects or Electrical Subjects) with the field of labor action where the women currently graduated from the career work; This result is similar to the research carried out at ECLAC in Santiago, Chile, where it is pointed out that the results of this study confirm that academic performance is a necessary condition for a successful insertion in the labor field. [20]

6. Conclusions

It is concluded that the percentage of women who graduate in relation to the percentage of men, by academic semester 2012-II to 2018-II of the career of Mechanical and Electrical Engineering of the National Technological University of South Lima, is equal to 8.1%.

It is concluded that the weighted average of the academic performance of women graduated in the professional career of Mechanical and Electrical Engineering, is 12.98, this represents a high average for the career; so it can be indicated that the 38 women graduated from the Mechanical and Electrical Engineering career have had a good academic performance.

It is concluded that there is a significant high ratio of 0.822 between the weighted average by specialty with the field of labor action; These results agree with what has been pointed out in relation to the fact that there is a higher qualification in the

specialty of Mechanics (AM) with an average range of 11.73 to 15.42, and according to professional practice reports, of the 38 women graduated from The career of Mechanical and Electrical Engineering, 25 are developing in the field of specialty of Mechanics.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors wish to recognize and thank the National Technological University of Lima South Lima for their support of this investigation.

References

[1] Tobal, S. Factors that influence the decision of engineering choice as a career in women. Latin American University of Science and Technology, 1-35. (2016).
[2] Panaia, M. The inclusion of women in the engineering profession. In: Virajes Magazine, Vol. 16, No. 1. Manizales: University of Caldas. 19-43 (2016).
[3] Sardina, M. Women and Engineering. Gender influence in career choice. University of cantabria. España. (2017).
[4] Gutiérrez, S. & Duarte, M. Engineers or engineers: how women are conceived in the engineering field. Autonomous University of Baja California. 89-97 (2017).
[5] Martínez, K. Professions have sex. Men and women in female and male professions, the case of nurses and electrical mechanical engineers. University of San Luis. San Luis Potosi (2016).
[6] Berlien, V.; Variela, P. & Robayo, C. I study national reality in training and promotion of women scientists in science, technology, engineering and mathematics. Ministry of Education. Chile (2017).
[7] Avolio, B.; Chávez, J.; Vílchez, C. & Pezo, G. Factores que influyen en el ingreso, participación y desarrollo de las mujeres en carreras vinculadas a la ciencia, tecnología e innovación. Pontificia Universidad Católica del Perú (2018).
[8] Vega, D. Women in Electrical Mechanical Engineering. University of Piura. UDEP. 102-110 (2017).
[9] Marsh, V. (2017). Stereotypes affect the academic performance of women. Journal of Science, Technology, Society and Culture, 85-93.
[10] Castro, M. The research project and its elaboration scheme. Caracas: Uyapal. (2017).
[11] Hernández, R. Investigation Methodology Mexico: Mcgraw Hill (2018).
[12] Arias, F. Research project Caracas: Episteme (2016).
[13] Pedroza, H. & Dicovskyi, L. Statistical Analysis System with SPSS. Nicaragua. INTA (2016).
[14] Rodríguez, M. Association Coefficients Mexico: Plaza and Valdez (2016).
[15] Ramos, R. Development and validation of instruments. University of Granada. Melilla (2017).
[16] Cárdenas, M., Eagly, A., Salgado, E., Goode, W., Heller, L., Jauregui, K., Galarza Quirós, N., Gormaz, N., Bunse, S., Godoy , M., Rocha Sánchez, T., Navarro, M., Sosa, F., Aguilera, Y., Schulmeyer, M., Tanure, B., Naranjo, M., Soto, B., Darre, S. and Tunqui, R. "Latin American female business executives: an interesting surprise", Gender in Management, Vol. 29 No. 1, pp. 2-24 (2014).
[17] Capital Peru: only 25% of women study engineering careers due to gender bias. Capital. Peru (2019).
[18] Álvarez. F.; Arias, A .; Serrallé, J. & Varela, M. Choice of engineering studies: Influence of scientific education and gender stereotypes on student self-esteem. Journal of Research in Education, No. 12 (1), pp. 54-72 (2016).
[19] Braginsk, R. A key profession for the future of the country. Austral University. Clarion (2019).
[20] Garbanzo, G. Factors associated with academic performance in university students, a reflection from the quality of public higher education. Education Magazine, 31 (1) (2007).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 373-381 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

360-Degree Videos, VR Experiences and the Application of Education 4.0 Technologies in Malaysia for Exposure and Immersion

Airil Haimi Mohd Adnan[*,1], Mohamad Syafiq Ya Shak[1], Rafidah Abd Karim[2], Mohd Haniff Mohd Tahir[1], Dianna Suzieanna M Shah[1]

[1]*Universiti Teknologi MARA Perak Branch (UiTM), Seri Iskandar Campus, Perak, 32610, Malaysia*

[2]*Universiti Teknologi MARA Perak Branch (UiTM), Tapah Campus, Perak, 35400, Malaysia*

ARTICLE INFO

Article history:
Received: 01 January, 2020
Accepted: 28 January, 2020
Online: 07 February, 2020

Keywords:
360-degree videos
Content development
Education 4.0
Immersive learning
Learning technologies
Virtual Reality

ABSTRACT

Obstacles, both imagined and real, continue to hinder the wider adoption of Education 4.0 learning technologies although these technologies are available in the mainstream consumer market. At the same time, the boom in Industry 4.0 manufacturing has brought down the prices of these technological tools making them generally more affordable. A case in point is 360-degree spherical video cameras and software that can record, render and playback immersive 'real life' contents. 360-degree spherical video contents can then be rendered and post-processed into VR (virtual reality) experiences that are not just immersive but also allow for limitless exposure time to learners. Both technologies allow for learner-driven learning to happen in the truest sense, for instance for second and foreign language learners. This research paper examines data collected from 560 undergraduates (n = 560) who were immersed and exposed to 360-degree videos and VR experiences for degree level courses on English (as a second language), Mandarin (Chinese) (as a third or foreign language, and Arabic (as a third or foreign language) at four public university campuses in Peninsular Malaysia. The benefits of 360-degree videos and VR experiences in second and foreign languages content delivery were evident to support learner-driven learning: Contents were developed by local lecturers and then uploaded for free online and learners could learn anytime, anywhere; total immersion could be achieved using cheap VR goggles powered by learners' smartphones; and most beneficially, weaker learners who desperately need more time to understand and practice difficult degree level second or foreign languages now have the freedom to revise and upskill themselves at their own pace.

1. Introduction

The leading futurist Gerd Leonhard [1] predicts that human civilisation will see more changes in the next 20 years than we have ever experienced in the last 300. If this prediction comes true, then those who are able to 'see' these changes before they happen and those who are responding in an active manner will have a lead over others who are merely reacting to those changes after they have happened. Indeed, there are changes happening now that are revolutionising the jobs that we do and the ways that we learn. Driven by technical developments never witnessed before, the nature of modern living as we know it will never be the same [2].

Within the realm of formal education, rapid changes are also happening. We have reached a critical juncture in teaching and learning because of Industry 4.0 disruptions [3, 4]. Within this context, educators and learners need to change the way they think and perceive the world from "*Why should we be doing this?*" to "*Why aren't we doing this?*" [5, 6]. One of the momentous changes happening right now within education is the convergence of technologically-enhanced immersive learning with learner exposure and learner-driven learning. These three concepts might be different, but they relate to the same key purpose which is to help our learners to learn in a more engaging, productive and useful manner compared to just learning by rote and memorising disjointed facts and figures that are easily forgotten [7]. Within the wider umbrella of immersive learning initiatives, learners at all stages of formal education in developed and developing nations are being exposed to interactive environments both physically and virtually; these are environments that can reproduce real world

[*]Corresponding Author: Dr. Airil Haimi Mohd Adnan, admin@uforia.edu.my

scenarios to teach certain skills. Combining simulations, game elements and virtual environments, 'true' immersive learning is finally made possible with technological developments from Industry 4.0 and supported by the philosophies of the Education 4.0 global movement [8].

2. Review of Research Literature

In this section, research literature that relates to the empirical inquiry reported in this paper is reviewed starting with an in-depth discussion of the notion of immersive learning followed by 360-degree videos and virtual reality (VR) experiences within the realm of education. The discussion is framed within the broader picture of the Industry 4.0 era and Education 4.0 universal movement.

2.1. Immersion, exposure and learner-driven learning

Within the realms of education and training, new and affordable Industry 4.0 technologies are creating new ways to make immersive learning more experiential and learner driven whilst opening opportunities to 'transport' learners into controlled environments with engaging tasks or situations for them to engage in. Figure 1 depicts the movement from less interactive learning materials to more immersive learning materials. The direct result of this is the ability of learners to learn at such a great depth than previously thought possible. As learning technologies start to quickly bridge the gap between what is conceivable and what is practical, deeply immersive learning is now feasible and it will continue to deliver more memorable and meaningful learning 'experiences' in future [1].

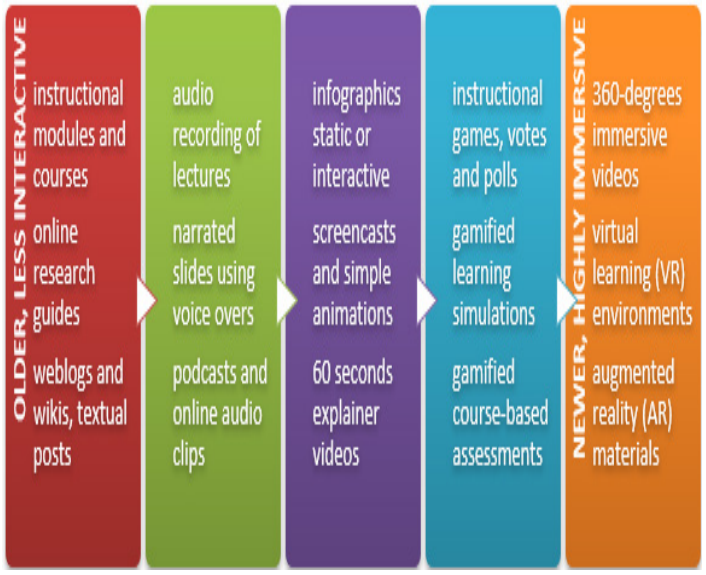


Figure 1: Digital learning objects from older and less interactive to newer and highly immersive ones

At tertiary level, to create an immersive learning experience traditionally might be limited to the course instructor telling a story or role playing a situation with learners [9]. For university or college instructors with more resources and time, instructional videos or online learning games offer some sort of immersion by drawing on experiences that emulate real-life situations and enable practice to happen in real time. Indeed, research suggests that experiences gained during learning simulations can be stored in our memory as if they were real [10]. At the same time, there are limitations with these approaches [11, 12].

Traditional flat 2D video clips and online learning games for example are limited in their learning value. Plus, they are far from immersive though they might be somewhat interactive. Worse, some learners can finish watching them or playing them quickly, lessening exposure time that is desperately needed by weaker learners to learn and relearn skills that might be difficult for them. At the other end of the spectrum, 'true' immersion that combines the best of e-Learning with engaging simulations is highly complex to produce and will therefore come at a higher price tag. Unsurprisingly, even for universities and colleges with a lot of cash, the price factor is always a burden in learning content development and deployment, not just in developing countries but in developed ones as well [13].

2.2. Learning through 360-degree videos and virtual reality (VR) experiences

Thankfully, rapid technical and technological developments of Industry 4.0 manufacturing have led to the creation of a new generation of immersive learning technologies that can be fully utilised by university or college instructors [14]. These include interactive 360-degree video technology, virtual reality or VR experiences, augmented reality or AR overlays, and mixed reality (MR) or extended reality (XR) simulations. Together, these technological tools hold exciting possibilities for bringing an immersive and learner-driven learning experience into the lecture room.

360-degree videos enable learners to observe a scene in whichever direction they wish. Thus, they are able to virtually explore a made up world or view an actual recording of the real world made by 360-degree video cameras. On smartphones, as learners move and turn their devices left and right or up and down, the images that they see move in perfect synchrony; on laptops computers and desktops, they can easily navigate spherical 360-degree videos by clicking and dragging navigation buttons. YouTube, for example, is witnessing a great number of 360-degree videos being uploaded not just for entertainment but also for teaching and learning purposes. On newer platforms like VeeR (which markets itself as the 'YouTube for VR') and using more advanced commercially-available software, 360-degree videos can be further rendered and post-processed into standalone 'virtual reality experiences'. VR experiences represent the cutting edge of technologically-enhanced teaching and learning, and they can be powered by standalone VR headsets capable of immersing and exposing learners to new learning experiences [15, 16].

With reference to VR experiences, whether they are fully computer-generated environments or built upon 360-degree videos from real-life situations, these experiences lead to immersive learning and longer exposure time for learners without real life difficulties, even though the initial setup will need investments to defray high costs. VR experiences enable learners to learn first-hand from their mistakes. And, it allows for longer exposure to learning materials and more practice time, especially for weaker learners who might have difficulties to acquire specific skills and knowledge from the classroom time alone. Immersing learners in real-life situations using interactive VR experiences is especially useful to train and develop soft skills, second language and also

third language skills [17]. Other skills like customer services or people management and formal meeting or negotiations are great examples where learners can be placed in positions of potential tension but where they are able to 'see' how certain decisions will lead to certain outcomes, without risks.

By learning about workplace-related scenarios in an interactive VR environment, for instance, the learner will be able to develop appropriate workplace-related skills to resolve conflicts, dilemmas, and problems in an appropriate manner. In other words, this really is learning by doing in the truest sense of the term. This is due to the fact that we have reached an exhilarating point in the rise of immersive learning technologies, as several factors converge to transform the future of learning landscape [18]. VR technology has now evolved from a personal computer tethered technology to one that can be easily accessed using smartphones and tablets. At this moment in time, an immersive VR experience is attainable merely by using the same gadgets that we carry in our pockets and bags. Furthermore, thanks to current VR platforms like VeeR, there is no need for special apps to be installed on smartphones. Add to this the wider affordability of cheap and easy to use smartphone-powered VR headsets (compared to 5-10 years ago), the investment needed to access immersive VR experiences is becoming lower all the time for learners at all levels of formal education.

2.3. 360-degree videos and virtual reality (VR) experiences for Education 4.0

The global Education 4.0 movement should be viewed as a set of trends and challenges that tertiary educators need to be aware of and to respond to (see Times Higher Education, 2019). The first trend is the transformation of the teaching and learning process. Tertiary educators must reconsider why and how they teach once Artificial Intelligence (AI) and other deep learning technologies become more prevalent in colleges and universities. The second trend is more personalised learner-driven learning. Tertiary educators must be able to cater to the learning styles of students and consider their individual behaviours, differences and performances. The next trend is personalised assessment. As AI systems run classroom tests, experiential learning through digital technologies will become the norm, so tertiary educators need to do without the high-stakes pen and paper tests. The final trend is the growth in intelligent digital environments. As digital environments become more common in physical classrooms, tertiary students need better experiences to be able to interact effectively and learn from and within those environments [19].

Today, the technological tools, software and apps related to Education 4.0 and Industry 4.0 that are needed to produce VR experiences are becoming more affordable and more accessible with newer and cheaper 360-degree cameras arriving on the commercial market [20]. Fully digital VR worlds could also be designed and built, albeit with technical know-how and a lot of time at hand. Consequently, the stage is set for immersive learning that is fully driven by learners to enter the learning ecosystem in formal education [21, 22]. Figure 2 illustrates VR technologies in education that are currently available for teaching and learning.

Still, there are issues and problems to contend with before these disruptive Education 4.0 technologies could spread from technically-savvy content developers to mainstream adoption [23]. The question is not whether something is new, but rather will it be better than what we have had in the past? [24, 25]. This is where hard evidence needs to be collected and analysed to truly examine whether immersive VR experiences are better for teaching and learning, in the tertiary education sector especially, compared to the teaching and learning methods that we have had in the past like 'chalk and talk' and rote learning that are supported by regimented drills and repetition exercises [26]. To answer these and other pertinent questions related to immersion, exposure and learner-driven learning through 360-degree videos and VR experiences, an empirical inquiry was carried out as explained in the next section.

Figure 2: VR technologies that are currently available in education

3. Research Methods and Research Participants

This study examines quantitative and qualitative data collected from 560 undergraduates (N = 560) who were immersed in, and exposed to, 360-degree videos and VR experiences for several degree level courses on English Language (second language), Mandarin Chinese (third/foreign language), and Arabic Language (third/foreign language) at four public university campuses in northern Peninsular Malaysia. 357 of the undergraduates are female and the other 203 are male; 136 of them were doing first degrees in art and design, and other architecture, planning or surveying-related courses, whereas the other 424 were first degree students in science and technology courses including engineering, computer sciences, applied sciences and other related courses. Overall, the data collection and analysis process took about ten months to complete, spanning two semesters, from the start of 2019. In particular, the study sets out to find answers to three research questions (RQ):

RQ1: What are the advantages and disadvantages of 360-degree videos and VR experiences in delivering educational materials, specifically for tertiary level second and third (foreign) language courses?

RQ2: Compared to methods used by 'traditional' instructors, how useful are 360-degree videos and VR experiences, for the current generation of undergraduates who are studying second and third (foreign) language courses?

RQ3: In the future, will 360-degree videos and VR experiences become more common as part of immersive and interactive learning initiatives, not just for the teaching of second and third (foreign) language courses? Why?

3.1. Research instruments and research cycles

In Cycle One of the studies, quantitative data collection was done using an online survey questionnaire with 25 items that asked the respondents to compare between 360-degree video materials and VR experiences with traditional learning in terms of their usefulness, usability and overall impression. They were also asked to respond to items related to the future of 360-degree videos and VR experiences, and the strength and weaknesses of these technologies with regards to language teaching and learning (both for second and third or foreign languages). Additionally, they were asked to respond to some survey items that are related to the contribution of 360-degree videos and VR experiences to their abilities and skills in either the English Language, Mandarin Chinese, or Arabic Language.

This was followed by Cycle Two that involved the collection of qualitative data employing focus group discussion sessions. These sessions were tied in with post-evaluation sessions during the actual class time to address the actual performances of the students involved in the three language courses. Five open questions were posed to a group of four to five learners regarding the same subject matters covered in Cycle One. The students (participants) were free to share their own ideas, feelings and opinions regarding the teaching and learning of either the English Language, Mandarin Chinese, or Arabic Language employing the 360-degree videos and VR experiences.

3.2. Data management and data analysis

The data collected from Cycle One were analysed based on mean, mode and median frequencies. The analysis of the data was done to show the prevalence of certain criteria based on the experiences of the respondents, for instance: the usefulness and usability of 360-degree videos and VR experiences; the perceived future of 360-degree videos and VR experiences; the strengths and weaknesses of these technologies with reference to second and third language education; and the influence of 360-degree videos and VR experiences on the abilities and skills of the respondents with regards to specific English Language, Mandarin Chinese, and/or Arabic Language skills.

The data collected from Cycle Two were more substantial and 'thick' [27] given the fact that the participants were asked to share their ideas, feelings and opinions in an open manner amongst their peers. Other than selective transcriptions of the focus group data, the participants were also requested to write a summary of the points and ideas that they talked about during the session. These were then checked and reshared with the participants to allow for 'member checking' [28] to happen. No clear-cut protocols were prepared for the discussion sessions and only open questions were posed to allow the participants to make comments and share their experiences freely. Sessions were conducted fully in either English or Malay language, as the national language of Malaysia (particularly for the Mandarin Chinese and Arabic Language undergraduate students).

After transcribing and coding, the data were thematically analysed in two stages: horizontal (group data) and vertical (personal data). At the end of the data collection and analysis stage, the three research questions were answered satisfactorily.

4. Research Results and Discussion

The results of this study are divided into two subsections as seen below: Cycle One and Cycle Two. The data for Cycle One are numerical/quantitative whereas the data for Cycle Two are textual/qualitative. Salient data are presented in this section and critically discussed, in an attempt to address three research/guiding questions in this empirical effort.

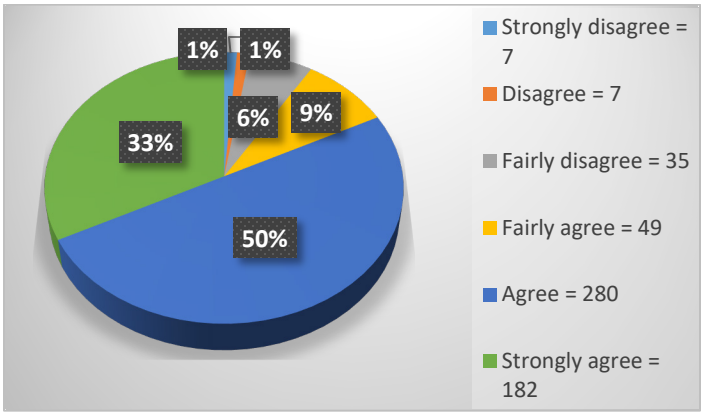
4.1. Cycle 1: Numerical/quantitative data

560 undergraduates (i.e., survey respondents) completed a 25-item survey questionnaire online where they were asked to compare 360-degree video materials and VR experiences with more traditional learning methods. They were also asked to think about the future of 360-degree video materials and VR experiences in tertiary education, and to gauge how these learning technologies have influenced their specific second or third language abilities and skills (in the English Language, Mandarin Chinese, or Arabic Language).

Using a Likert scale, the respondents were able to choose between 'Strongly disagree', 'Disagree' or 'Fairly disagree' at the negative end of the spectrum whilst at the positive end they could choose either 'Fairly agree', 'Agree' or 'Strongly agree'. Neutral or unsure responses were not included as these are typical distractors that student-respondents will normally choose, based on our personal experience in doing such research.

The first item in the online survey questionnaire asks the respondents whether the use of 360-degree videos and VR experiences felt more useful to them compared to more traditional teaching methods like chalk and talk, drilling and also lecturing. Perhaps unsurprisingly and given the fact that the respondents have all been exposed to 360-degree videos and VR experiences to some degree beforehand, 511 or 91.25% of them chose positive answers, as seen in Graph 1 below.

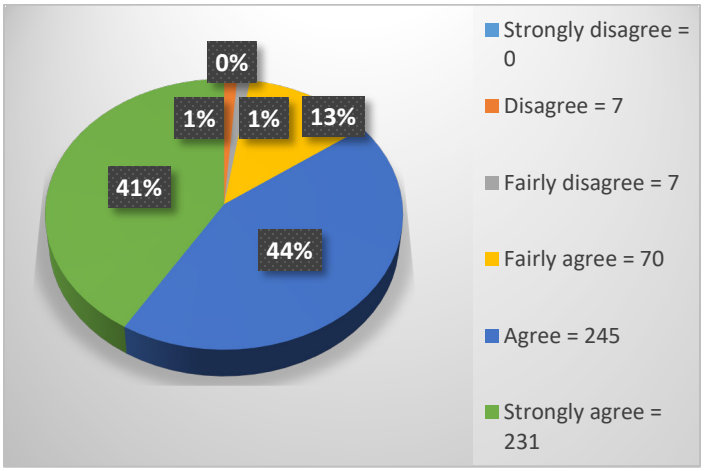
Graph 1: "Using 360-degree videos and VR are more useful than traditional learning methods" (from 560 survey respondents, n = 560)



The greatest strengths of 360-degree videos and virtual reality (VR) experiences lie in their abilities not just to make the learning more immersive and interactive, but also to allow learners or end-users to gain access to limitless exposure time to specific learning materials, be it for the teaching of languages or any other tertiary level subjects. Indeed, 360-degree videos and VR allow learners to learn as much as they want, when they want to; the more time a learner spends being immersed in these technologies, the more

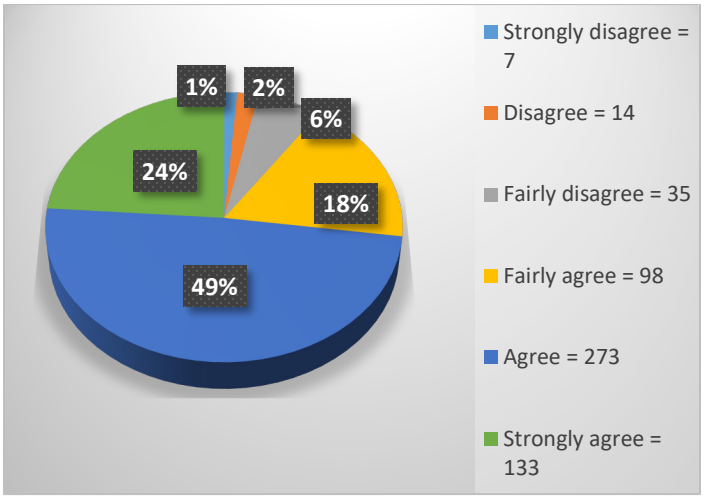
prepared she or he will be for formal assessments and other tests related to the language courses her or she is taking. This fact is not lost on the respondents as shown in Graph 2.

Graph 2: "Using 360-degree videos and VR I can get more time to practice before facing tests" (from 560 survey respondents, n = 560)



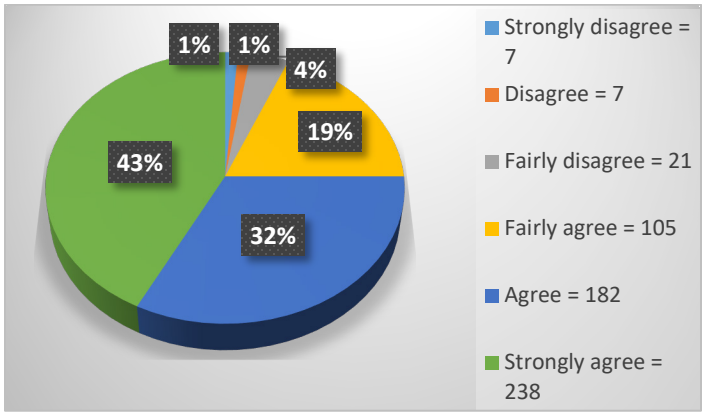
Graph 2 above is a clear indicator of the usefulness and utility of 360-degree videos and also VR experiences in terms of helping learners to practice intensively and extensively before they face course assessments or tests. However, bear in mind that these statements relate mainly to the teaching and learning of English Language, Mandarin Chinese, and Arabic Language. And so, it is quite surprising to find that the respondents were equally open to the idea of using 360-degree videos and VR experiences for the teaching and learning of other subjects at university level, as shown in Graph 3 below.

Graph 3: "360-degree videos and VR experiences should be introduced for other university subjects" (from 560 survey respondents, n = 560)



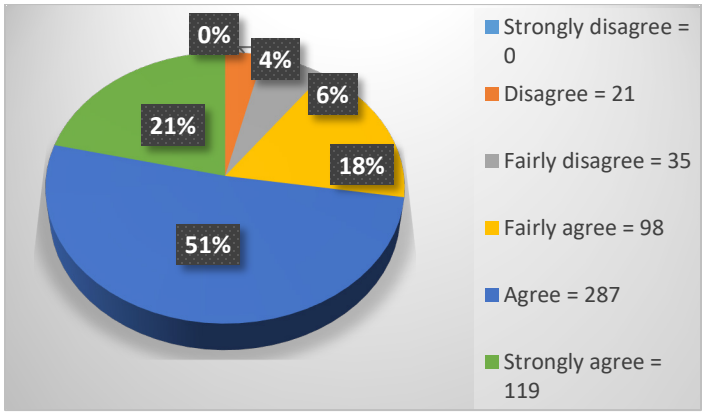
Interestingly, the respondents also showed very positive orientation towards the use of 360-degree videos and VR experiences at university level not just for individual subjects or courses. They believe that the administration of the universities should apply these technologies for future teaching and learning initiatives, so that other students too can benefit from immersive and interactive learning technologies. The responses recorded are presented in Graph 4, next.

Graph 4: "In future, my university should use more technologies like 360-degree videos and VR" (from 560 survey respondents, n = 560)



In another survey item related to Graph 4, the respondents were asked whether their lecturers should be more open to the idea of using 360-degree videos and VR experiences in the teaching and learning process as shown in Graph 5. This is because one of the biggest obstacles in the adoption of 21st century learning technologies is the disinterestedness of tertiary lecturers and course instructors. Perhaps these lecturers and instructors feel that traditional methods are still needed or maybe they think that teaching and learning with technology is an unneeded hassle within their campuses. Without a doubt, these data must be used to inform future empirical studies that examine the roles of tertiary lecturers and instructors as content developers for future learning initiatives.

Graph 5: "The lecturers at this university need to be open to 360-degree videos and VR experiences" (from 560 survey respondents, n = 560)

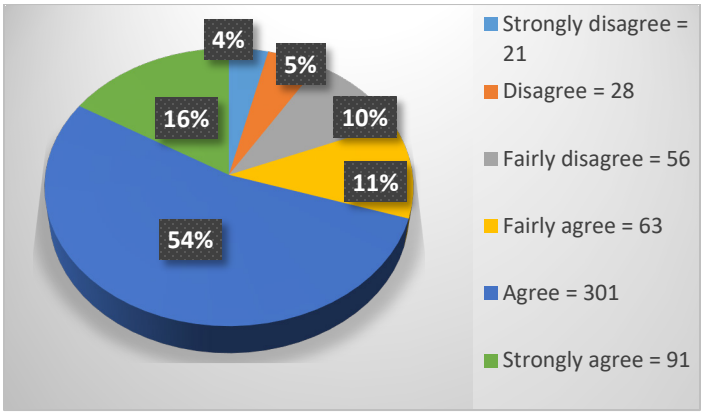


Items 15 and 16 in the survey questionnaire are related to the technical side of the deployment of 360-degree videos and VR experiences in the teaching and learning dyad. Even though learners can interact with 360-degree videos using YouTube, in the form of spherical or 360-degree video objects, to truly become immersed in the learning process they need to invest in VR goggles. There are many types of VR goggles that can be procured but the cheapest and most accessible, especially for university students, seem to be the VR goggles that are powered by their own smartphones. Users need only mount their smartphones within the goggles to access interactive and immersive VR experiences.

Higher immersion leads to a better engagement with the subject matter at hand, in this case the English Language, Mandarin

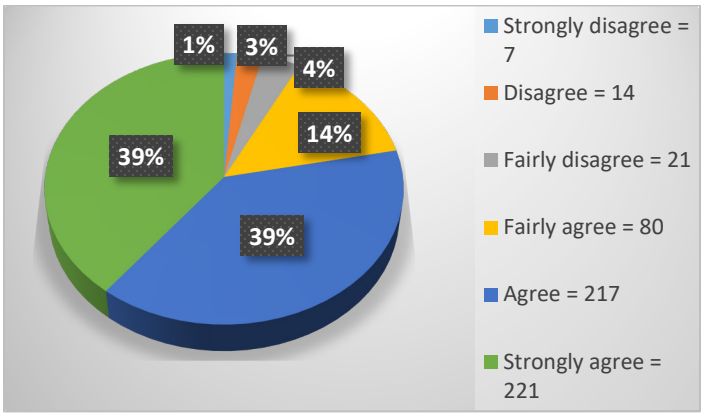
Chinese, or Arabic Language. Higher immersion and the ability to interact with VR experiences also lead to more exposure time which can help with the acquisition and practice of specific skills. Graph 6 shows that many of the respondents were interested in buying VR goggles as part of their learning process at university.

Graph 6: "I think VR goggles are a good investment if used for teaching and learning at university" (from 560 survey respondents, n = 560)



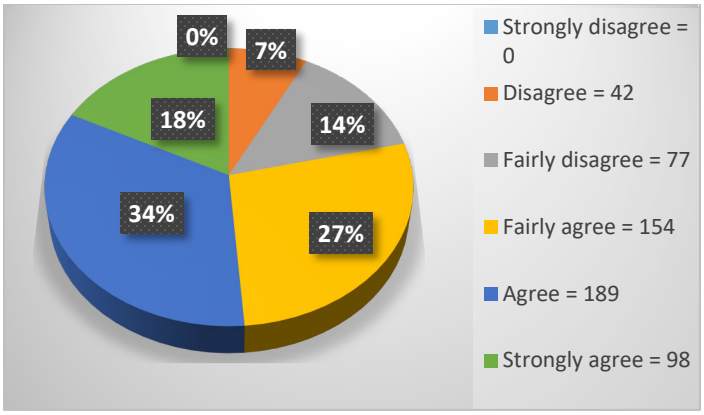
It must also be stated that although the respondents generally harbour positive views on the use of 360-degree videos and VR experiences in the teaching and learning process, there are several technical obstacles that hinder the wider adoption of these teaching and learning technologies. Understandably, it must be frustrating for university students to have to contend with mobile data limits and bad Internet connectivity when they want to use 360-degree videos and VR experiences to learn new things and to practice new skills. That being said, these annoying issues seem to plague and to hinder a wider use of Internet-related teaching and learning technologies at this present moment, in Malaysia and in most developing countries. Graph 7 portrays the frustrations of the respondents.

Graph 7: "360-degree videos and VR experiences are cool, but the problem is that our Internet sucks" (from 560 survey respondents, n = 560)



The final graph, Graph 8, is a clear indication that Malaysia as a developing nation is moving in the right direction in terms of deploying 360-degree videos and VR experiences in the teaching and learning process at university level. Only 119 respondents or 21.25% of the population under study have somewhat negative feelings in terms of the contribution of these technologies to their skills and knowledge (in the English Language, Mandarin Chinese, or Arabic Language) compared to 441 respondents or 78.75% who have accepted these technologies in a positive light. In other words, for every five undergraduates at university, four will benefit from the use of 360-degree videos and VR experiences in the teaching and learning process.

Graph 8: "Contribution to learning (VR 360-degree videos and VR experiences) to my level of skill/knowledge at the END of my course in either English or Mandarin or Arabic languages" (from 560 survey respondents, n = 560)



4.2. Cycle 2: Textual/qualitative data

Cycle Two of the data collection and analysis process relied on focus group discussions to gain insights into the ideas, feelings and opinions of the participants with regards to the use of 360-degree videos and VR. These focus group discussions were linked with the post-evaluation of test performances for the students involved in the English Language, Mandarin Chinese, or Arabic Language courses.

Five open questions were then posed to a group of four to five participants regarding the same subject matters covered in Cycle One. The names below are pseudonyms chosen by the participants to protect their identities but the comments shared are actual comments and utterances with only minor editing to improve accuracy and to follow academic writing conventions.

4.2.1. Benefits and advantages of VR experiences and 360-degree videos at university

According to 'Siti Noorul Nadiah', "Technology that uses 3-dimensional graphics makes us, the users, feel like we are really interacting in the real world. We can also visualise abstract learning materials, so these materials become easier to understand, more enjoyable and fun to learn." Siti's ideas and feelings seem to be shared by nearly all the participants, due to the fact that they have all been exposed to VR experiences and 360-degree videos.

For 'Nurul Ainal Ainaa,' the biggest benefit of using VR experiences and 360-degree videos in the teaching and learning process is the exposure to new technologies that are relevant in the era of Industry 4.0. She further explained:

> Technology like this is really quite expensive to buy or develop. So, when we use this in class, students can have the opportunity to learn using the state-of-the-art VR technology and 360-degree videos. My friends and I also noticed that the learning process can be a lot more interesting and fun by using VR and 360-degree videos, but the problem right now is that our lecturer doesn't have much materials to teach us with this advanced technology.

4.2.2. Problems faced when using VR experiences and 360-degree videos at university

Echoing Nurul's last sentence, 'Che Mamat' said, "I feel there should be many more things (contents) using VR and 360-degree videos but I also understand they're not easy to create. Maybe if there are more courses using VR, more students can learn using this technology." Another challenge for university students when they use technologically-enhanced learning materials is related to technical issues. Problems linked to limited access to mobile Internet data and stability issues when they use public networks on campus, commonly lead to audio loss, video stutter and other technical annoyance that lessen the overall quality of their immersive VR experience. In all fairness, these issues are not just annoying, but they could hinder the wider acceptance and use of technologies like VR experiences and 360-degree videos in the long run. In the words of 'Ahmad Afnan':

> Other than the connection problems, the most common problem at this time is that not all students have VR goggles, so it's quite complicated to watch the videos because I need to make the phone stable to focus at the best point, and I need to scroll the screen to see the whole video all the time. This takes a lot of my time and sometimes I lose the key points in the VR videos. Next, the sound is slow. Although, I am using earphones with the highest volume I cannot hear clearly… Our lecturer must create VR with higher quality in future.

4.2.3. Feelings after being chosen to test VR experiences and 360-degree videos for the different second or third (foreign) language courses

On the whole, all of the participants were appreciative in being given the chance to test and use VR experiences and 360-degree videos for their English Language, Mandarin Chinese, or Arabic Language courses. They particularly love the opportunity to try out the technology, the chance to get more exposure to key skills in their course and the chance to learn in a modern way that is totally different from the traditional methods that they are used to. As 'Siti Zaleha' observed:

> This semester I was exposed to new technology and I learned about the professional worker of the future. For English, in class, we don't just learn about English, but the lecturer also teaches about how to use 360° Degrees (videos) and Virtual Reality out of class using our own initiatives. As students we will become more advanced in technology and I think this style is an effective learning method, similar to the 21st century learning concept. I think we need more chances to learn in this method even for the other subjects and courses.

4.2.4. Improvements needed to make VR experiences and 360-degree videos more useful

Many participants commented about the overall 'production value' of the 360-degree virtual reality experiences that needs to be improved, if the project is planned for a wider dissemination and use in the future. They were particularly annoyed with the low sound quality of the VR experiences even though they feel that the videos are of sufficiently high quality. The fact that these VR experiences and 360-degree videos were prepared for English Language, Mandarin Chinese, or Arabic Language courses where communication skills are tantamount to success makes this a thoroughly valid point. As 'Zachary' mentions, "Another way is you should probably put subtitles to support the audio of the video because I couldn't really hear well. Also, everyone's pronunciations are different so subtitles will come in handy." Another enlightening comment was made by 'Nurain Ahmed':

> Maybe the video doesn't have to be related to lectures all the time? I mean maybe, the lecturer can also do online question and answer sessions from time to time or maybe we play online question games, so that students can enjoy the class more. The online games and online quizzes can test their grammar or vocabulary for the professional workplace, maybe? Students will definitely feel more excited to do these things because the class will feel more relaxed. Then, we can just revise on our own using VR after the class time.

4.2.5. Whether the participants support (or not) a wider deployment of VR and 360-degree videos

From the 560 respondents / participants involved in this study, 553 or an impressive 98.75% supports the use of VR experiences and 360-degree videos for teaching and learning at university. Just seven participants or 1.25% does not support the wider deployment of these Education 4.0 learning technologies. For 'Siti Zafirah', she is against the use of these technologies for two reasons. First, the network access to the Internet on campus is very bad. And second, she feels that she has no choice but to buy the VR goggles. 'Ahmad El-Zickry' also disagrees with any initiatives to widen the use of VR and 360-degree videos:

> I strongly disagree with VR-360 technology because it requires strong Internet connection. However, as you know in this campus the Internet connection is very weak and too slow for learning. When I use the VR and 360-degree videos, it turns out to be someone else talking but someone else's face had come out. What kind of learning is this? This is rubbish!

For the majority of the respondents, they believe that VR experiences and 360-degree videos represent the future of teaching and learning in Malaysia. They, therefore, support these technological initiatives even though they too face some difficulties related to Internet connection and not having access to VR goggles. 'Alia Ibni Bakar' shared her insights on this matter:

> I agree totally with the usage of the VR and 360-degree videos. It's time to change our teaching and learning style. This is a new road that our university must take if they want to use this teaching style. Some of us students, we also want to experience new things while we're still studying. So, I think this new method of learning and teaching is good for students like myself. Maybe by using this method, students will feel more excited to learn difficult subjects and it will be easy for them to revise subjects? All we have to do is just replay and replay stuff.

5. Conclusion

There *are* barriers to the wider adoption of Education 4.0 learning technologies at university level in Malaysia [21]. These barriers are mostly related to technical issues, like patchy connections to the Internet and limited materials available that take advantage of the VR experiences and 360-degree videos in delivering quality educational contents. That being said, these

barriers should *not* stop Malaysian university lecturers and instructors from becoming learning technologists who strive to make learner-driven learning a reality. In truth, technological tools, development software and related apps are now easily available in the mainstream consumer market that allow for the creation and development of cutting-edge teaching and learning materials. In addition, the boom in Industry 4.0 manufacturing has brought down the prices of these tools, software and apps, making them more affordable for self-funded university educators to start their own teaching and learning innovation projects.

This empirical study contributes to the literature on education, and teaching and learning for year 2020 and beyond, looking specifically at VR experiences and 360-degree videos to deliver quality educational contents to teach specific (and difficult) skills in the English Language, Mandarin Chinese, and Arabic Language courses at four different public university campuses in a developing country. Employing 360-degree spherical video cameras and software that can record, render and playback immersive real-life contents, these contents can then be post-processed into VR experiences that are not just immersive but offer limitless opportunities for learning exposure, as depicted in Figure 1 at the start of this empirical paper.

Both technologies allow for learner-driven learning to happen, for a group of 560 second and some third (foreign) learners at four different public university campuses in Malaysia. The results are extremely encouraging, as seen in the numerical and textual data records presented in this paper. Immersion followed by further repeated exposure has helped the learners (end-users) to acquire the specific skills that they need and to practice for course-related assessments even with limited classroom contact hours. With these concrete data, it is possible to imagine that VR experiences, 360-degree videos and other Education 4.0 teaching and learning technologies will soon move into the mainstream and bring true the learner-driven 21st century learning to university classrooms in Malaysia and in other developing nations around the globe.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The corresponding author would like to thank his colleagues at all research sites for their willingness to try out 360-degrees video technology and virtual reality or VR technologies that he had developed for the teaching and learning of second and third (foreign) language courses (at their respective university campuses in Peninsular Malaysia). The corresponding author also acknowledges the assistance of Mr. Muhamad Khairul, Mr. Muhammad Anwar, Miss Nurul Nadiah and Mr. Ahmad Ariffuddin, young short-term contract lecturers from his Future Learning Initiatives Lab (FLI LAB @ UiTM), in helping to create and develop the 360-degrees video experiences. Special thanks to Mr. Mohamad Syafiq as the 'forensic proof reader' of this paper. It should be noted that this empirical project received no funding neither from the authors' institutions nor from the government of Malaysia; any and all costs incurred in the process of carrying out this academic project were borne by the authors/researchers ourselves.

References

[1] G. Leonhard, "As technology becomes cognitive, all paths must lead to collective human flourishing." https://www.forbes.com/sites/cognitiveworld/2019/02/25/as-technology-becomes-cognitive-all-paths-must-lead-to-collective-human-flourishing/, 2019.

[2] A. Davies, D. Fidler, M. Gorbis, Future Work Skills 2020. Palo Alto, CA, Institute for the Future, University of Phoenix Research Institute, 2011.

[3] K. Schwab, The Fourth Industrial Revolution. Geneva, World Economic Forum, 2016.

[4] K. Schwab, N. Davis, Shaping the Fourth Industrial Revolution. Geneva, World Economic Forum. 2018.

[5] World Economic Forum, The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution. Geneva, World Economic Forum, 2016.

[6] World Economic Forum, The Future of Jobs Report 2018. Geneva: World Economic Forum, 2018.

[7] U. D. Ehlers, S. A. Kellermann, Future Skills: The Future of Learning and Higher Education. Karlsruhe, International Future Skills Delphi Survey, 2019.

[8] A. H. M. Adnan, "Industry 4.0 skill sets, higher-order thinking skills and gamification: Lessons from 'Potentia Project', Malaysia" (Plenary Panel Presentation). Proceedings of TING X 2018: The Tenth National Meeting of Teachers 'Innovation in Education for Indonesia 4.0'. Tangerang: Universitas Terbuka Indonesia, 2018.

[9] AdvancED, "Learning in 3D: Making STEM Real." https://www.advanc-ed.org/source/learning-3d-making-stem-real, 2015.

[10] 3DLabs (2019). "Virtual Reality and Augmented Reality." http://www.3d-labs.co.uk/virtual-reality/, 2019.

[11] A. H. M. Adnan, M. K. Ahmad, M. A. Mohd Kamal, N. N. Mustafa Kamal, A. M. Yusof, N. M. Azamri, "Virtual Reality Immersion for 'Education 4.0' English Education: ELSA 360°-Videos Project – Phase 2." In MNNF Publisher (Ed.), Leading Towards Creativity & Innovation (Series 2) (pp. 329-343). Senawang, MNNF Publisher, 2019.

[12] A. H. M. Adnan, M. K. Ahmad, A. A. Yusof, M. A. Mohd Kamal, N. N. Mustafa Kamal, "English Language Simulations Augmented with 360-degree spherical videos (ELSA 360°-Videos): 'Virtual Reality' real life learning!" In MNNF Publisher (Ed.), Leading Towards Creativity & Innovation (Series 1) (pp. 82-88). Senawang, MNNF Publisher, 2019.

[13] J. O. Wallgrün, A. Masrur, J. Zhao, A. Taylor, E. Knapp, J. S. K. Chang, A. Klippel, "Low-Cost VR Applications to Experience Real Word Places Anytime, Anywhere, and with Anyone." In 2019 IEEE 5th Workshop on Everyday Virtual Reality (WEVR) (pp. 1-6). IEEE, 2019.

[14] A. Doucet, J. Evers, E. Guerra, N. Lopez, M. Soskil, K. Timmers, Teaching in the Fourth Industrial Revolution: Standing at the Precipice. Oxford, Routledge, 2018.

[15] M. K. Ahmad, A. H. M. Adnan, A. A. Yusof, M. A. Mohd Kamal, N. N. Mustafa Kamal, "Using new technologies to teach English in Malaysia – issues and challenges." In MNNF Network (Ed.), Proceedings of the Int'l Invention, Innovative & Creative Conference, Series 1/2019 (pp. 203-207). Senawang: MNNF Network.

[16] Aniwaa, "The Ultimate VR, AR, MR Guide." https://www.aniwaa.com/guide/vr-ar/ultimate-vr-ar-mr-guide/, 2019.

[17] K. N. Mohd, A. H. M. Adnan, A. A. Yusof, M. K. Ahmad, M. A. Mohd Kamal, "Teaching Arabic Language to Malaysian University Students using Education Technologies based on Education 4.0 Principles." In MNNF Network (Ed.), Proceedings of the Int'l Invention, Innovative & Creative Conference, Series 2/2019 (pp. 38-51). Senawang, MNNF Network, 2019.

[18] M. Rüfenacht, "Education 5.0 why we need to adjust the education system." https://medium.com/@MattiaSuisse/education-5-0-why-i-think-we-need-to-adjust-the-education-system-4a669 b26396d, 2017.

[19] N. N. Mustafa Kamal, A. H. M. Adnan, A. A. Yusof, M. K. Ahmad, M. A. Mohd Kamal, "'Immersive Interactive Educational Experiences' – adopting Education 5.0, Industry 4.0 learning technologies for Malaysian universities." In MNNF Network (Ed.), Proceedings of the Int'l Invention, Innovative & Creative Conference, Series 1/2019 (pp. 190-196). Senawang, MNNF Network, 2019.

[20] P. Madigan, "A spotlight on virtual reality in the English classroom." mETAphor, 2, 56-58, 2018.

[21] A. H. M. Adnan, R. A. Karim, M. H. M. Tahir, N. N. Mustafa Kamal, A. M. Yusof, "Education 4.0 Technologies, Industry 4.0 Skills and the Teaching of English in Malaysian Tertiary Education." Arab World English Journal, 10(4), 330-343, 2019. DOI: https://dx.doi.org/10.24093/awej/vol10no4.24.

[22] R. A. Karim, A. G. B. Abu, A. H. M. Adnan, A. D. J. Suhandoko, "The Use of Mobile Technology in Promoting Education 4.0 for Higher Education." Advanced Journal of Technical and Vocational Education, 2(3), 34-39, 2018. DOI: https://dx.doi.org/10.26666/rmp.ajtve.2018.3.6.

[23] J. Taplin, Move Fast and Break Things: How Facebook, Google, and Amazon Have Cornered Culture and What it Means for All of Us. London, Pan Books, 2017.
[24] Times Higher Education, "Preparing for Education 4.0." https://www.timeshighereducation.com/hub/jisc/p/preparing-education-40, 2019.
[25] A. H. M. Adnan, Z. M. Zamari, "'I am a Techno-Rebel!' Malaysian Academics & their Personal Experiences of Progressing into e-Learning." Procedia-Social and Behavioral Sciences, 67, 61-72, 2012. DOI: https://doi.org/10.1016/j.sbspro.2012.11.307.
[26] A. H. M. Adnan, "The use of, and attitudes toward, the English Language in Bumiputera Malay boarding schools in Malaysia: Reality of the Malaysian 'English language dilemma,'" Master of Arts Thesis, University of Leicester UK, 2001.
[27] C. Geertz, The Interpretation of Cultures: Selected Essays. New York, NY, Basic Books, 1973.
[28] C. A. Chapelle, P. A. Duff, "Some guidelines for conducting quantitative and qualitative research in TESOL." TESOL Quarterly, 37(1), 157-178, 2003.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 382-386 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Optimization of the Feeding Point Location of Rectangular Microstrip Patch Antenna

Md. Biplob Hossain[1], Md. Selim Hossain[2], Md. Mahabub Hossain[3,*], Md. Dulal Haque[3]

[1]*Department of Electrical and Electronic Engineering, Khwaja Yunus Ali University, Sirajganj, 6751, Bangladesh.*

[2]*Department of Computer Science and Engineering, Khwaja Yunus Ali University, Sirajganj, 6751, Bangladesh.*

[3]*Department of Electronics and Communication Engineering, Hajee Mohammad Danesh Science and Technology University (HSTU), Dinajpur, 5200, Bangladesh.*

ARTICLE INFO

Article history:
Received: 23 October, 2019
Accepted: 23 January, 2020
Online: 20 February, 2020

Keywords:
Microstrip Patch Antenna
Substrate Material
Feed Point
Pattern of Radiation
Directivity

ABSTRACT

Microstrip patch antennas (MPAs) are comfortable and have better anticipation compared to traditional antennas. Not only the weight of this type of antennas is lighter but also, they have little capacity, minimum price, reduced in measurement as well as affluence of fiction then observance. Recently, it has been many times reevaluated for the improvement of Rectangular Microstrip Patch Antenna (RMPA). The performance of an antenna relies on dimensions of antenna, operating frequency, substrate material and height of substrate, the thickness of patch as well as different feeding techniques of the antenna. The main concept that highlights this work is to determine the finest location of the feed point that plays a vital role in the performance measurement of the designed antenna and investigate the performance by studying distinct feed point location. Here we used Rogers R03006 (lossy) as substrate material whose dielectric constant is 6.15 and the operating frequency of our designed RMPA is 1.22 GHz. The RMPA is analyzed based on the different constraints factors of antenna like loss of return, voltage standing wave ratio (VSWR), and the improvement factors that mean the directivity, gain, overall efficiency, the pattern of radiation, etc. It is revealed that the best feeding point (9.07, 15.75) of our designed RMPA is unique concerning the substrate materials and antenna height, as well as width, at a specific operating frequency.

1. Introduction

Definitely, it's a progressive field in the direction of plan an MPA and numerous designs have been anticipated in current years [1]. It is notified, wavelength plays a vital role to determine the size of an antenna and its size is equivalent to the wavelength as well as inversely proportional to the operating frequency. MPA comprises of a patch, substrate, ground plane, etc. The microstrip antenna may have square [2], rectangular [3], circular [4] and elliptical [5] shapes, but any continuous shape is possible. The constant of dielectric (ε_r) value of the substrate should be kept within the values of 2.2 to 12 and for the other materials, it should be as low as possible while designing a microstrip patch antenna [6,7]. Loss tangent reduces efficiency and affects antenna performance. From top to bottom gain, directivity, and competent efficiency of radiation are obtained when the low loss tangent materials are used.

Diverse simulation-based software has become favored recently for designing a patch antenna [8]. We have used CST (CST STUDIO SUITE 2017) simulation software for designing the MPA which is commercially recognized as a simulation-based software to design microstrip patch antenna. To determine the perfect feed point location for reducing the significant return loss is the main motive of our research work. Rogers R03006 (lossy) with dielectric constant ε_r of 6.15 as substrate material of our designed RMPA has been investigated at the operating frequency of 1.22 GHz. The loss of power input and reflected power is called the return loss which can be calculated in dB from the variance of the feeding and reflected power [9]. The applications of this antenna would be in global positioning system (GPS), wireless local area networks (WLANs), synthetic aperture radar (SAR), medical diagnostic operation in two or more discrete bands and dual-band antenna designs.

*Corresponding Author: Md. Mahabub Hossain, Department of Electronics and Communication Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur, 5200, Bangladesh.
Email: im.mahabub@gmail.com

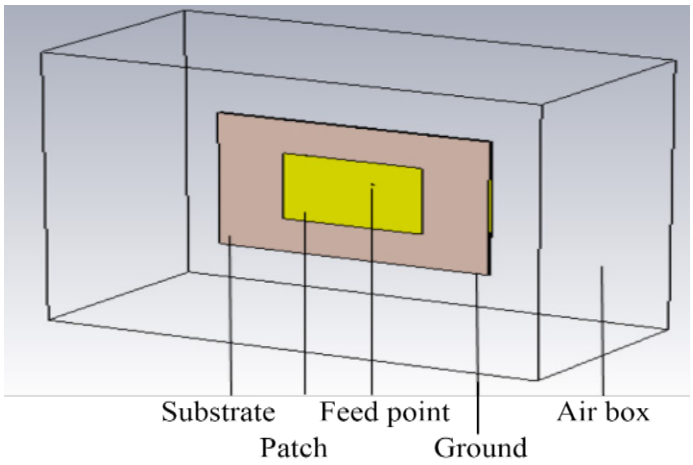


Figure 1: Perspective View of Rectangular Microstrip Patch Antenna in CST simulation platform.

2. Design Procedure of Rectangular Microstrip Patch Antenna

A RMPA with a simple form of design, as shown in Figure 1, comprises of radiating patch (RP) forming one side with a dielectric substrate (DS) and the opposite side with a ground plane (GP). On the dielectric substrate material, the radiating patch include with the feeder line is liked as photographs compose. The fringing field in the boundary of patch and ground plane of the microstrip antenna can emit radiation. The coaxial probe feeding method involves a central conductor that connects directly to the patch for providing radiating power [10].

It was used as a substrate material called Rogers R03006 (lossy) whose dielectric constants $\mathcal{E}_r$=6.15, permittivity (μ) =1 F/m, thermal conductivity=0.61 to design the proposed antenna. Here different operating frequency was taken between 1 to 2 GHz for analysis [9, 11].

$$W = \frac{C}{2fo}\sqrt{\frac{2}{1+\varepsilon_r}} \quad (1)$$

where W = patch width, fo= frequency of resonance and c = light speed.

$$L = L_{eff} - 2\Delta L \quad (2)$$

where, L = patch length, L_{eff} = effective length, ΔL =extension of length.

$$L_{eff} = \frac{C}{2fo}\sqrt{\frac{2}{\varepsilon_{eff}}} \quad (3)$$

where $\mathcal{E}_{eff}$ = substrate's effective dielectric constant.

$$X_f = \sqrt{\frac{L}{\varepsilon_{eff}}} \quad and \quad Y_f = \frac{W}{2} \quad (4)$$

where X_f = input position as feed point along X axis and Y_f = input location as feed point along Y axis.

$$\Delta L = 0.412h\frac{(\varepsilon_{eff}+0.3)(\frac{w}{h}+.264)}{(\varepsilon_{eff}-0.258)(\frac{w}{h}+.8)} \quad (5)$$

where ΔL=length extension.

$$\varepsilon_{eff} = \frac{\varepsilon_r+1}{2} + \frac{\varepsilon_r-1}{2}[1+12\frac{h}{w}]^{1/2} \quad (6)$$

where $\mathcal{E}_{eff}$ = substrate's effective dielectric constant.

$$L_g = 6h+L \quad (7)$$

$$W_g = 6h+W \quad (8)$$

The parameters values have been given below Table 1, which have been calculated from the above equations.

Table 1: Parameters and Values of Designed RMPA.

Parameters	Values (mm)
Patch width (W)	49.40
Patch length (L)	36.10
Ground plane width (W_g)	61.40
Ground plane length (L_g)	48.10
The patch thickness (t)	0.05
Substrate height (h)	2
Pin inner radius (r_i)	0.5
Pin outer radius (r_o)	2.5

3. Identifying the Feed Point Location

The performance of RMPA greatly depends on the location of feed point, which can be positioned at the point of (X_f, Y_f) where X_f and Y_f are the input feed point position along the x-axis and y-axis respectively. To locate the perfect feed point that relates to the better performance of RMPA, we have simulated in CST by taking values along x-axis 0, 5, 9.07,15,20 and y-axis 0, 5, 10, 15.75, 20. The theory-based feeding point location is calculated using the values of the parameters which are mentioned in Table 1 and we find the theory-based values along with (x, y) axis (9.07, 15.75) by applying the above equation that has been described previously. We have studied by taking some location-points above and below of (9.07, 15.75). Comparing the obtained values of loss of return, VSWR, directivity, gain, overall efficiency, pattern of radiation, etc. from the CST simulated software, we got the best result in (9.07, 15.75) axis.

4. Results and Discussion

The theory-based feeding point location is calculated using the values of the parameters which are mentioned in Table1.

Table 2 indicates return loss of diverse feeding point location along with the x and y coordinates. From this table, it is found the minimum return loss which is -33.831 dB for the axis (9.07, 15.75). The best feed point location has been indicated with yellow color. We have exercised some values above and below of (9.07, 15.75) which show less performance compared with the best feeding point

location. In [12], the authors showed minimum return loss of -42.18 dB for the axis point of (-3, -3), when they used $LaAlO_3$ material a dielectric constant of 23.5, and a height of 1.5 mm as a substrate for the operating frequency of 2.4 GHz. In such a way, it can be shown that the best feed point location of an antenna depends on substrate material along with its heights as well as the operating frequency.

Table 2: Return Loss (dB) of 25 feeding point location in the designed RMPA.

Y- axis \ X- axis	0	5	9.07	15	20
0	-4.13	-4.13	-23.03	-4.54	-7.65
5	-4.73	-4.76	-15.17	-5.09	-8.80
10	-7.63	-7.57	-22.74	-7.47	-8.92
15.75	-29.04	-30.71	-33.83	-32.83	-32.1
20	-4.17	-4.27	-12.39	-5.11	-8.17

The CST simulated figure according to Table 2: is shown below with its necessary parameters and values.

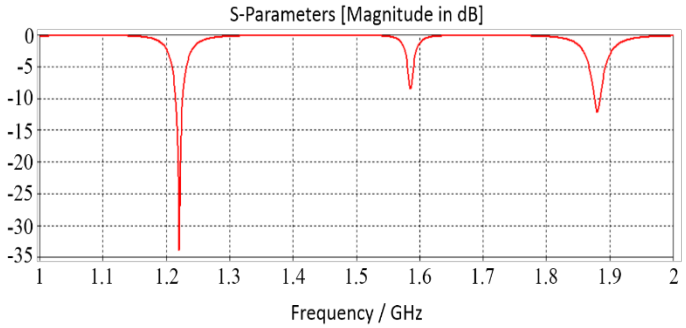


Figure 2 shows the S-Parameters value with respect to the frequency where feed point (9.07, 15.75). Here high return loss is -33.831 dB at frequency 1.22 GHz. A simulated result for different feeding point positions has been shown in Figure 3.

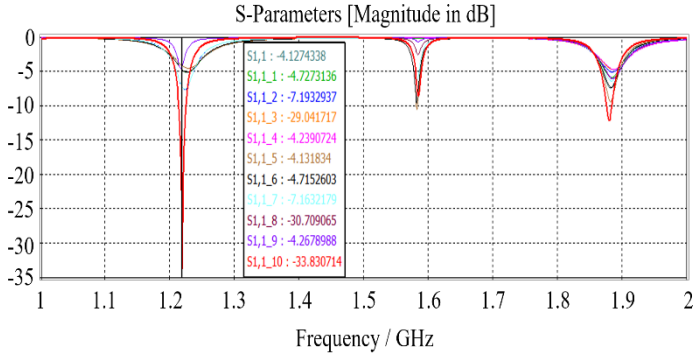


Figure 3: Simulated result for different feeding point locations.

We have examined 25 feeding point locations. In this graph, it has been compared to 11 results among the 25 feeding point locations. In our research work, we have found minimum return loss (-33.83 dB) as our desired feed point compares to the other feed point locations.

Table 3: Voltage Standing Wave Ratio (VSWR) of 25 feeding point location.

Y- axis \ X-axis	0	5	9.07	15	20
0	5.670	1.180	1.152	3.906	2.410
5	4.930	1.996	1.423	3.510	2.139
10	2.421	2.438	1.157	2.466	2.115
15.75	1.049	1.060	1.042	1.046	1.051
20	4.117	3.008	1.632	3.494	2.070

Table 3 indicates voltage standing wave ratio (VSWR) of diverse feeding point locations along with the x and y coordinates. From this table, it is observed the low VSWR which is 1.042 for the axis (9.07, 15.75). Here, the lowest VSWR value has been indicated by yellow color.

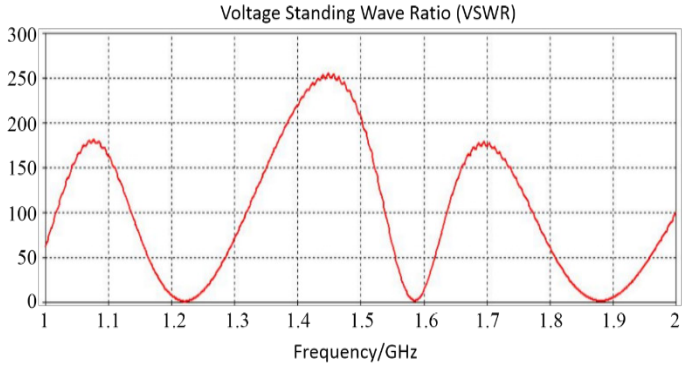


Figure 4: VSWR vs. Frequency at feed point (9.07, 15.75).

The CST simulated figure according to values of Table 3 is shown in Figure 4, which illustrates the VSWR with respect to different frequencies. It is observed the low VSWR which is 1.042 at frequency 1.22 GHz.

Table 4 demonstrates the directivity (dBi) of various feeding point locations along with the x and y coordinates. The directivity (dBi) for the axis (9.07, 15.75) is 6.44 dBi, which is very close to the maximum value of 6.48 dBi for the feed point of (20, 5), as shown in yellow color.

Table 4: Directivity (dBi) of 25 feeding point location in the designed RMPA.

Y- axis \ X-axis	0	5	9.07	15	20
0	6.44	5.17	6.41	6.43	6.47
5	6.42	6.40	6.39	6.41	6.48
10	6.42	6.42	4.56	6.41	6.44
15.75	6.41	6.41	6.44	6.40	6.40
20	6.40	6.41	6.41	4.55	4.45

Figure 5 expresses the directivity (dBi) at feed point (9.0, 15.75). It shows that the directivity is 6.44 dBi, which is very near to the maximum value of 6.48 dBi.

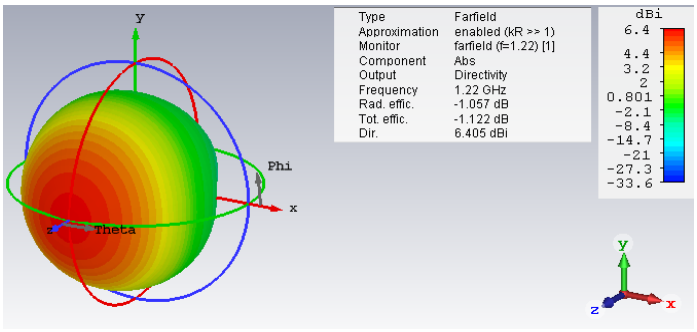


Figure 5: Directivity (dBi) at feed point (9.07, 15.75).

Table 5: Gain (dB) of 25 feeding point location in the designed RMPA.

Y-axis \ X-axis	0	5	9.07	15	20
0	5.37	3.23	4.74	5.37	4.81
5	5.35	5.31	4.71	5.35	4.84
10	5.35	5.37	2.65	5.35	4.72
15.75	5.34	5.35	5.35	5.35	5.34
20	5.32	4.75	4.73	3.11	2.07

Table 5 expresses the gain (dB) of various feeding point locations along with the x and y coordinates. From this table, we find the gain (dB) which is 5.35 dB for the axis (9.07, 15.75) which is very close to the maximum value of 5.37 dB. This feed point location has been indicated with a yellow color.

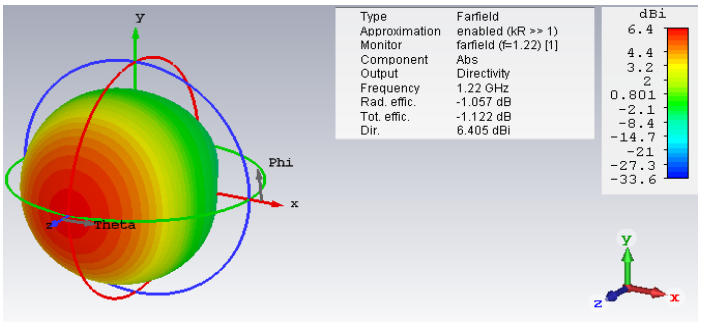


Figure 6: Gain (dB) at feed point (9.07, 15.75).

Figure 6 specifies the gain (dB) where the feed point is (9.07, 15.75). The gain is found 5.35 dB, which is near close to the maximum value of 5.37 dB. The energy radiation by an antenna can be shown by radiation pattern. A diagrammatical illustration of the radiation pattern for the indication of circulation of radiated energy into space is a function of direction. A 2D radiation pattern has been given in Figure 7 with its necessary values.

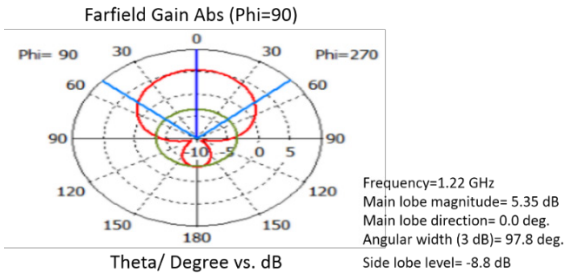


Figure 7: 2D radiation pattern of farfield gain at feed point (9.07, 15.75).

Figure 7 displays the radiation pattern at frequency 1.22 GHz where the feed point is (9.07, 15.75). From the pattern, it is found that the magnitude of the main lobe is 5.35 dB, whereas the side lobe magnitude is -8.8 dB. In addition to the 3D representation, we have added the polar diagrams in the three interesting planes which are farfield H-field, farfield E-field and farfield E-pattern.

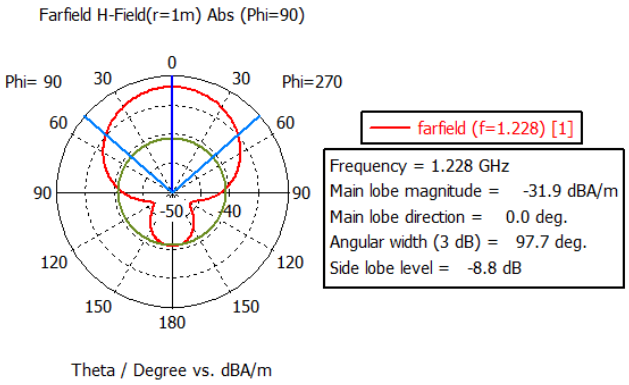


Figure 8: Farfield H-Field at feed point (9.07, 15.75).

Figure 8 indicates the farfield H-field at frequency 1.228 GHz where the magnitude of the sidelobe level is -8.8 dB. The same result for the side lobe has been observed for the farfield E-field and farfield E-pattern as shown in the Figure 9 and Figure 10 respectively.

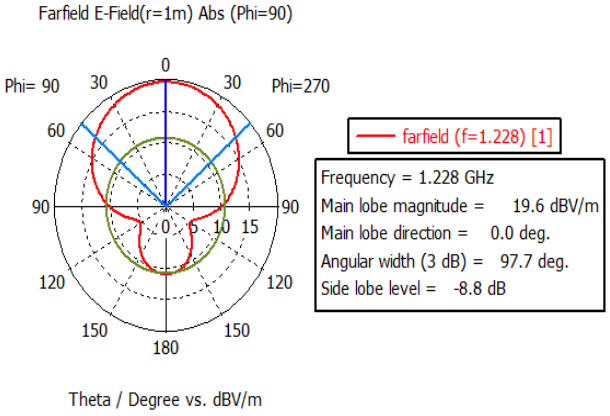


Figure 9: Farfield E-field at feed point (9.07, 15.75).

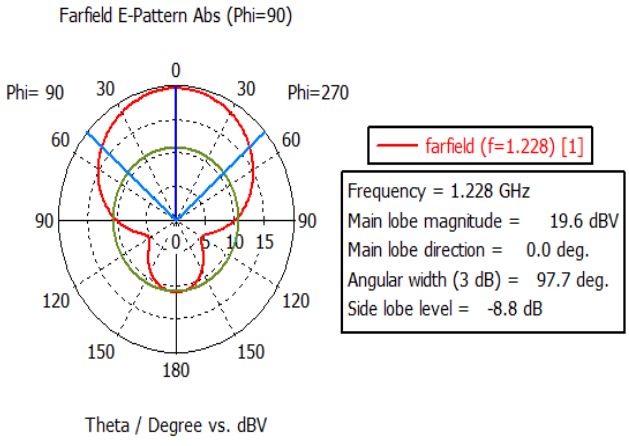


Figure 10: Farfield E-pattern at feed point (9.07, 15.75).

The ratio of the entire power released by an antenna to the net power enter to the antenna from the associated transmitter source is called radiation efficiency. In addition, total radiation efficiency includes the power lost due to poor VSWR or mismatch loss. From the Figure 11, it has been found that the radiation efficiency is -1.0565 dB, whereas -1.222 dB is the total efficiency.

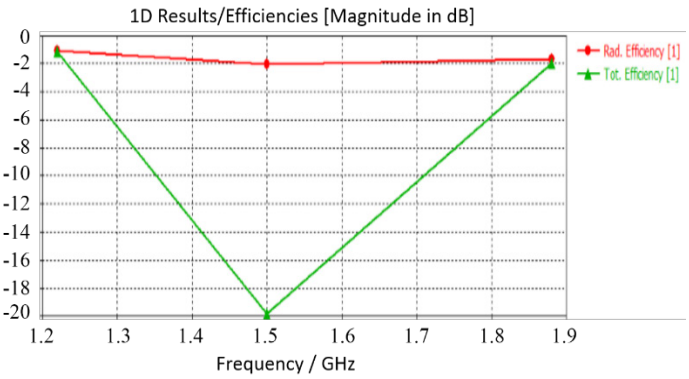


Figure 11. Radiation and total efficiency at feed point (9.07, 15.75).

Figure 11 exhibits radiation and total efficiency at different frequencies. It is found that the radiation and total efficiencies of the prospective antenna are -1.0565 dB and -1.222 dB respectively.

5. Conclusion

The performance of a RMPA is controlled through changing the feed point settings. The optimal outcome of 25 nourishing points of the designed RMPA on Rogers R030006 (lossy) substrate has been investigated. All the feeding point locations result for return loss, VSWR, gain, and directivity have been compared. It is shown that the perfect selection of the feed point location can significantly improve the gain, directivity and efficiency performance of a microstrip patch antenna by optimum reduction of the return loss and other constraints factors.

References

[1] Y. Ahmed, Y. Hao and C. Parini, "A 31.5 GHz Patch Antenna Design for Medical Implants", International Journal of Antennas and Propagation, Volume 2008, Article ID 167980, 6 pages, doi:10.1155/2008/167980.

[2] N. Porchelvi, "A Design of Square Fractal Antenna with Microstrip Feed For Ultra Wideband Applications", International Conference on Electrical, Information and Communication Technologies (ICEICT), 2017.

[3] M. M. Olaimat, "Comparison between Rectangular and Circular Patch Antennas Array", International Journal of Computational Engineering Research (IJCER), ISSN (e): 2250 – 3005, Volume-06. Issue-09, 2016.

[4] S. Jain, R. Nema, "Review Paper for Circular Microstrip Patch Antenna", International Journal of Computer Technology and Electronics Engineering (IJCTEE), Volume 1, Issue 3, ISSN 2249-6343, 2009.

[5] V. Priyashman, M. F. Jamlos, H. Lago1, M. Jusoh, Z. A. Ahmad, M. A. Romli, M. N. Salimi, "Elliptical Shape Microstrip Patch Antenna without Dots" IEEE Symposium on Wireless Technology and Applications (ISWTA), September 23-26, 2012, Bandung, Indonesia, 2012.

[6] D. Rano and S. D. Gupta, "Design and Analysis of Microstrip Antenna Array in E & H plane for High Gain Applications", 2015 International Conference on Signal Processing and Communication (ICSC), Noida, India, 16-18, 2015.

[7] M. S. Nishamol, "Design and Development of Reconfigurable Compact Cross Patch Antenna for Switchable Polarization", Department of Electronics, CUSAT, India, 2012.

[8] S. E. Jasim, M. A. Jusoh, M. H. Mazwir and S. N. S. Mahmud, "Finding The Best Feeding Point Location Of Patch Antenna Using HFSS", ARPN Journal of Engineering and Applied Sciences, ISSN 1819-6608, Vol. 10, No. 23, 2015.

[9] T.S. Bird, "Definition and misuse of return loss", [report of the transactions editor-in chief Antennas and Propagation Magazine. IEEE. Vol. 51, No. 2, pp. 166-167, 2009

[10] A. Mehta, "Microstrip Antenna", International Journal of Scientific & Technology Research. Vol. 4, No. 3, pp. 54-57, 2015.

[11] R. Garg, P. Bhartia, I. Bahl and Ittipiboon, "A Microstrip Antenna Design Handbook", Artech House. Boston-London. pp. 1-17, 2001.

[12] S. E. Jasim, M. A. Jusoh, M. H. Mazwir and S. N. S. Mahmud," Finding the Best Feeding Point Location of Patch Antenna Using HFSS", ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 23, 2015.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 387-402 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Trust and Reputation Mechanisms in Vehicular Ad-Hoc Networks: A Systematic Review

Amit Kumar Tyagi[1,2,*], A. Mohan Krishna[3], Shaveta Malik[4], Meghna Manoj Nair[2], Sreenath Niladhuri[1]

[1]*Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry-605014, India*

[2]*School of Computing Science and Engineering, Vellore Institute of Technology, Chennai Campus, Chennai, 600127, Tamilnadu, India*

[3]*Department of Computer Science and Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India*

[4]*Department of Computer Science and Engineering, Terna Engineering College, Navi Mumbai, Maharashtra, India*

ARTICLE INFO

Article history:
Received: 23 October, 2019
Accepted: 25 January, 2020
Online: 20 February, 2020

Keywords:
Vehicular Ad hoc Networks (VANETs),
Vehicle User,
Trust and Reputation
Security
Privacy in Vehicles

ABSTRACT

An emerging trend has been observed in the Trust and Reputation (T & R) systems in field of decision-making support for a majority of the provisions propagated by the Internet. It is the extreme importance that peers (users) are able to trust each other and rely on them for file sharing and for services. This paper provides the reader an apprehensive and completely true information and details on a large number of the present conceptions, proposals, issues, and the key to those problems in VANETs and other fields to enhance the eminence of the data in transportation through a systematized literature review. Trust and reputation have also been discussed gravely in this paper. After the scrutinized analysis of more than 90 articles related to trust in a plethora of fields, extracted from few of the apt scientific sources ((i.e., SIEEE Computer Society, ACM Digital Library, Springer Link, Science Direct, and Wiley Online Library), and hence, succeeds to bring about the major hurdles and necessities for trust in real world and future research.

1. Trust and Reputation (T&R) – A Detailed Discussion (Introduction)

In the past decade, trust has received a great attention in the field of psychology, sociology, economics, political science, anthropology and recently in wireless networks. Mundane life find trust to be at the zenith of priority and everything goes by accordingly [1]. After all, when you run into a communal society, trust proves to be a foundation for desired decision making and efficient rating strategies. There is no universal definition for trust and reputation. In this work, authors defined Trust as "a subjective assessment of another's influence in terms of the extent of one's perceptions about the quality and significance of another's impact over one's outcomes in a given situation, such that one's expectation of, openness to, and inclination toward such influence provide a sense of control over the potential outcomes of the situation" [1]. Reputation is the opinionized version of an individual regarding another person or an object. In Oxford dictionary, reputation is said to be an opinion or belief held on someone about something.

*Corresponding Author: Amit Kumar Tyagi, amitkrtyagi025@gmail.com

Abdul Rahman et al. [2] define reputation as "an expectation about an agent's behavior based on information about its past behavior". Chang et al. [3] define reputation as "an aggregation of the recommendations from all of the third-party recommendations agents and their first, second and third hand opinions as well as the trustworthiness of the recommendation agent in giving correct recommendations to the trusting agent about the quality of the trusted agent". Trust have been invented an essential term in human-beings life, so it should be maintained in vehicle applications (for example, during carpooling, parking) to encourage the vehicle users to perform transactions with the other users over road network.

1.1. Trust

Trust is the old word and came into existence along with human being evolved on this earth. It plays a significant role in the survival of human beings. As we experience trust on daily basis, it is not an objective property but subjective to the degree of belief shown on a person, process or objects. The degree of trust varies over situation, person and opinion. It is not a blind guess or a game of chance, but it is blind guess based on the

knowledge and experience acquired over a period of time. The basic understanding of trust is list as follows:

a. The development of a bonding between two or more individuals aimed towards a particular target of action or goal is referred to as trust, such that they trust and rely on each other for the smooth and successful fulfilment of the action. The first specimen is the subject, while the other is the agent. Hence, we can see that subject, agent and action can be used to define trust.
b. Uncertainty and doubtfulness are ways to measure trust. Consider the following three situations: (a) The case when the subject has cent percent trust on the agent and strongly believes that the agent can and will fulfil the action. (b) The case when the subject highly despises the trust worthiness of the agent and hence there is no possibility of uncertainness here as well, but in a different perspective when compared to the previous case. (c) The case when the subject has an extremely vague idea about the agent, leading to the rise of large levels of uncertainty as he doesn't trust the agent at all.
c. Trust is not symmetric in most of the cases. The thin line of trust between two entities A and B does not have to facilitate the very same feeling for either of them.

In general, trust can be a personal experience between the partners based on the context, reputation and on recommendations. It is the presence of uncertainty and depends on the expected risk among the partners while interaction. Further, we recall some characteristics of trust as follows [1]:

- Binary, Directed Relation: Trust is a binary directed relation linking two entities. It is considered as the confidence of an entity called trustor towards another entity referred to as trustee.
- Asymmetry: If entity A trusts entity B, it does not imply that entity B trusts entity A. Trust is not necessarily reciprocal between a pair of entities.
- Contextual: Trust is considered in the context of particular actions which the target entity may perform.
- Subjectivity: Trust is the level of the subjective probability with which the trustor assesses that the trustee entity will perform a particular action.
- Non-transitivity: If entity A trusts entity B, and B trusts C, it does not necessarily imply that entity A trusts entity C.
- Composability: Trust relations can be constructed between not directly connected entities. Entity A may query for trust of entity B. Many entities in the community can provide different ratings, i.e., reputation scores of trusts for B. Thereafter, entity A can aggregate the received information to assign a trust level for B based on its own trust assessment method.
- Self-reinforcing: Entity tends to act honestly with trusted entities and to abstain with untrusted ones. Thus, this behavior reinforces the trust relation among trusted entities along the time.
- Dynamicity: The trust score of an entity may change over time. It may increase or decrease depending on the current performance of the entity. If its current interaction has a better quality than the last ones, its trust level could increase, and vice versa.

Properties of Reputation based Trust Systems are Complexity, Decentralization, Dynamics, Scalability, Privacy, Security level, Sparsity and Robustness, etc. A comparison among trust properties can find in table 4.

Trust Parameters: The certainty of agents being capable enough of doing an action at par excellence. This has a the two-tiered definition: Firstly, agents must develop the habit of believing their peers; secondly, its completely in the hands of the agent to delegate actions to the peer. In a social environment, trusty nature can often be defined as the facilities handed over to the other earlier irrespective of their relationship in the present. Content storage and exchange are the seemingly leading areas in the field of trust within P2P technology. With advancement in modern ways of living, most of the models, focus on malicious behaviors and capacity to complete the transactions.

The five factors for evaluation of trust are as follows:

- The feedback obtained between peers.
- The final and all in all dealings happening between the peers.
- The certainty of the source of feedback
- The framework for distinction of important transactions from the less important ones.
- The social factor for addressing the corresponding issues.

The word 'trust' describes that belief and expectation about future behavior based on past experiences and evidences collected either directly or indirectly.

1.1.1 Belief (or faith)

Belief is the essential and important to make trust and reputation among people. Trust is "a peer's belief in another peer's capabilities, honesty and reliability based on its own direct experiences". Reputation is "a peer's belief in another peer's capabilities, honesty and reliability based on recommendations received from other peers". The decision by agent X to delegate a take to agent Y is based on belief and this belief is called 'trust'. To build a mental state of trust, the basic beliefs that an agent need are:

- Motivation belief- X believes that Y has some motivation to help X and this motivation over long-term help X to achieve his goal. If y believes to be motivated, then x tends to trust him.
- Competence belief: The optimistic mindset of the agent to assure himself that the task can be accomplished by agent Y. Else, the feeling of trust would be of utter waste.
- Dependence belief: The agent believes that better to rely on y to complete the task successfully.
- Self-confidence belief: X strongly believes in y for completing his task.
- Disposition belief: Though it may not be necessary that Y could do the task, articulation of disposition belief and support would gather great help along with two more beliefs:
    - Willingness belief: The agent is assured about Y doing α (the action that leads to the goal g). X is confident about doing the required proposal as suggested by Y. If Y feels uninterested in doing the task, they might as well end up masking themselves saying that they intend to do so. However, this

would immensely decrease the bond of trust between them.

- Persistence belief: The agent acquires a positive approach and a touch of stability in Y about the task completion of *α.* If Y's stability seems to dwindle, there are chances of risks being produced when they interact with Y, due to which there's just mere belief existing between the two.

These two beliefs together refer to what is called soul trust and reliance. Along with these, the following also arises:

- Fulfilment belief: If the agent believes that Y has the potential to complete g, the agent would decide

i) not to drop the plan of the goal ii) Not to bring about changes or modifications to it iii) to encourage Y to complete it.

In short, we can prove that trust revolves around a mentalistic feature which would showcase the agents mind (X) who prefers the work to be done by Y, Y being the intellectual agent. So, in short, X is completely assured of Y doing the action and Y continues.

Kinds of Trust

Trust and faith are complementary to each other and revolve around the sense of confidence. The only thing that distinguishes the two is nothing but the fact that trust consists of a risk of element but confidence does not. We often notice trust to be used in common geek speak, when people have a spark of feeling inside them, i.e., an internal emotion that keeps oscillating, on the basis of the situation. Trust is also considered to have phenomenal feeling of trusting or being trusted. Figure 1 discusses the kind of trust in term of computation (refer figure 1 after Appendix A).

Basically trust (among human being) can be positive, negative, global, local or based on recommendation or reputation value. In other words, based on the reputation or recommendation, trust can be positive, negative, global or local. The contravention of trust leads to mental fluctuations between the agent and the source. Hence, on the grounds of creative design agents open systems, trust can be summarised as shown:

- Basic Trust: It is a general trust disposed independently on the agent based on the good experience accumulated by the agent.
- General Trust: Without taking any specific situation into consideration, one agent will trust another agent.
- Individual-level Trust: The agents has some belief on their partners
- System-level Trust: By the rules of the regulatory bodies in the system, the agents are forced to be trustworthy.

There seem to be three different forms of trust on the basis of a societal research. This includes dispositional trust [1, 4], system trust [5], etc. There are numerous perspectives and takes on trust and it's not practical to furnish each one of them. Several Type of Trust discussed in [4]. Further, Trust Classes discussed as: For specificity on trust semantics, distinction between a large number of different trust cases on the basis of Grandison and Sloman's classification (2000) [6] has also been analysed..

- Provision Trust: It talks about the trust values of the dependent party and is to be looked into when they seem to acquire coverage from harmful and disloyal service providers. Business trust [7] is commonly used to showcase the communal interactions between the companies which emerge from the contract agreements that keeps a track of the trust relations between them. For example, when a contract mentions a qualified delivery service, business trust would act aa a provision trust in our language.
- Access Trust: It talks about the trust with respect to a principiated dimension wherein the dependent party's resources are being targeted at. This relates to the centralised component of any computer which is the access control paradigm. Grandison & Sloman (2000) [6] provides an excellent perspective of access trust systems.
- Delegation Trust: It personifies trust as an agent (delegate) and decides on behalf of the dependent party. It has been well pointed out by Grandison and Sloman [6] that one of the specific service provisions would be to act on one's behalf.
- Identity Trust: It portrays belief to be a claimed identity. X.509 and PGP are the trust systems which produce their very own identity [8]. This is a topic of interest in the information security society, of whose overview can be found in in Reiter and Stubblebine (1997) [9].
- Context Trust: It tags the limit up to which any dependent party trusts the necessary systems and machinery to facilitate safe and sound transactions. These can be keenly affected by critical architectural framework, insurance, legal system, etc.
- Experience Trust: It refers to the trust that an agent has obtained on the previous and past interactions with a client. The interactions are called as transactions and the trust thus obtained is transaction trust.
- Similar Trust: This is the trust than an agent acquires by reasoning with respect to a client with other clients. A client is considered to be well known if there's an active conversation existing between the client and the agent and both of them have their own experience trust about each other.

Trust Models in VANET: The trustworthy nature of peers is the key focus in entity-oriented trust models. On the contrary, data-oriented trust models focus on the loyal nature of the peers. Data-oriented trust models depend on evaluating the trustworthiness of the transmitted data. In such models, there is no long-term trust relationships between nodes are formed. The combination of a number of trust models use node's trust to evaluate the aptness of data. Figure 2 defines this concept/ picture clearly (refer figure 2 after Appendix A). Basically, trust depends on friendship-not money, not power, not education. Only if there is trust, will there be friendship. Whenever we got friendship, we got feedbacks or recommendation for future to maintain our trust among other person. A Detail description about various trust establishment models existed in VANET network discussed in table 3.

Trust Calculation

If two nodes want to communicate each other, the level of trust worthiness of two nodes must be high enough to continue the communication process. Before the initialization of a connection, nodes refer to the local table and seek help from the adjacent

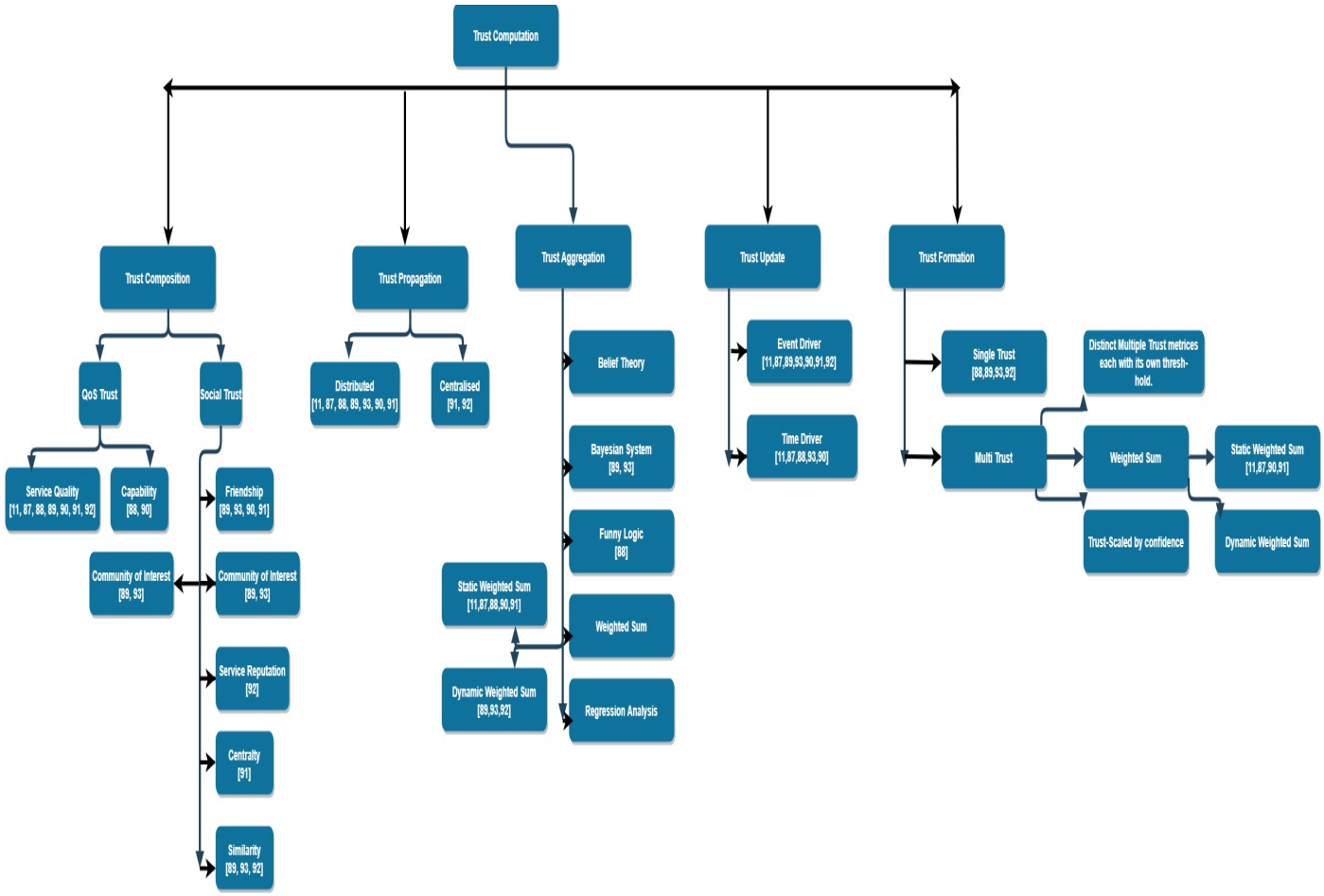


Figure 1: Classification Tree of Trust

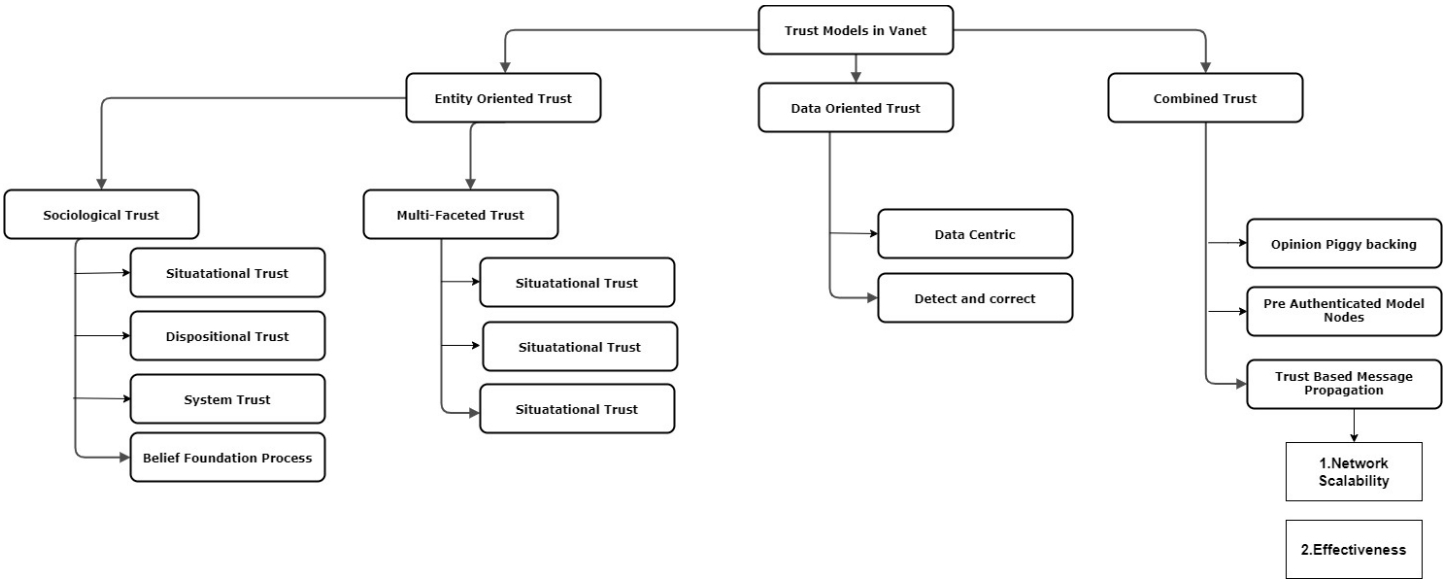


Figure 2: Trust Models in Vehicular Ad-hoc Network (VANET)

nodes to identify the trustworthy nature of the colleague. “Small World Problem” is a new approach to ease out and enhance the reach of the nodes to its neighboring nodes for trust evaluation. The decision made by every node is highly dependent on the prioritization given by them to the array and the minimum value too accepts them as per the “Trust Array” approach [10].

Some of the critical node accepts the communication request whereas few of them will reject it. Trust array values should be maintained somewhere on application level and recalculation of the values are required whenever the node initiate a new communication. Here trust model defines three important metrics: service, recommendation and reputation (see figure 3). Figure 3

represents variation of trust values when user starts a new interaction.

- Service trust metric is essential not only for evaluating the trustworthiness of the neighbor but also for the selection of the best service provider.
- Recommendation and reputation used to Measuring the trustworthiness of the new node. Recommendation trust metric include the number of recommendations for the stranger from various neighbor nodes.
- Reputation is required during the preliminary stages of communication. Later the number of acquaintances increases and hence the importance of reputation is less.

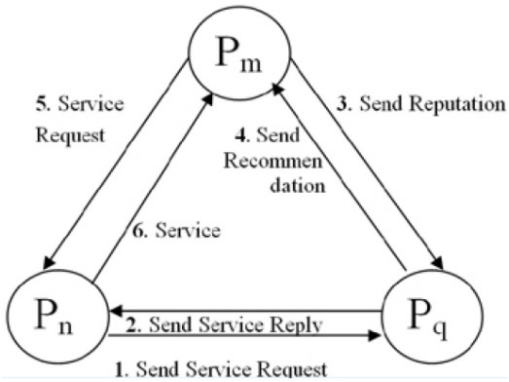


Figure 3: Interaction between the Peers

Computational Trust: Figure 4 shows that, Trust can be compute based on direct evidence (direct availability trust, direct integrity trust and direct competence trust) and indirect evidence (indirect availability trust, indirect integrity trust and indirect competence trust). The work on trust computations can be broadly classified into the following categories:

- Distributed trust computations: Each individual node creates its analysis on its peers through neighbor sensing, trust systems based on recommendation, and hybrid method.
- Centralized trust computations: It is a centralized form of computation and manipulation of the evaluations.

The central controls and manipulates the trust evaluations. In section 2, this work is provided in detail, i.e., advantages or disadvantages about each trust models (centralized, decentralized and partially decentralized). Through stringent sieving and combination of sophisticated technologies, around seven critical elements can be shortlisted in figure 4. Furthermore, through filtering combining the duplicated terminologies used in the different trust models, seven critical compositional elements can be summarized in figure 4. Further, dimension is to study what the sources of the trust values are:

- The syntax mainly focuses on the actual definition of trust if the outcome is a combined product.
- The model inculcated for the calculation of trust is called trust computation engine.
- Trust network includes the study on the pattern set up of the agents and the host agents.
- Uncertainty involves efficient risk management which supervises the occurrence of incidents and reliability which assures the reliability of the agent.

Values of trust are often classified into one or multi dimensions. A very specifically bound context limiting itself to the description of trust is found in a single dimensional approach. On the contrary, a multidimensional approach inculcates a feeling of unassured trust. The framework of trust values revolves around: rating, ranking, probability, belief, fuzzy value, etc. The rating is often related to numerical numbers of natural type. For example, on a scale of [11, 1], 1 is used for representing “Highly Untrustable”, while 4 is linked to “Highly Trustable”. Discrete tag-names are used for the labelling of the levels of trust, namely “Very Trustworthy”, “Trustworthy”, “Untrustworthy”, and “Very Untrustworthy” for direct which is mainly used for direct trust while “Very Good”, “Good”, “Bad”, and “Very Bad” are used for the recommenders’ trust. Local and Aggregate are the two types of rating [12]. Ratings based on the personal interactions with the second peer is called Local Rating and is produced whenever an interaction takes place. Ratings which are computed with the local ratings and validations from witnesses as a base is called Aggregate Rating. This is mainly used for computing if the peer is faithful or loyal and if they are worthy enough to be propagated to other peers.

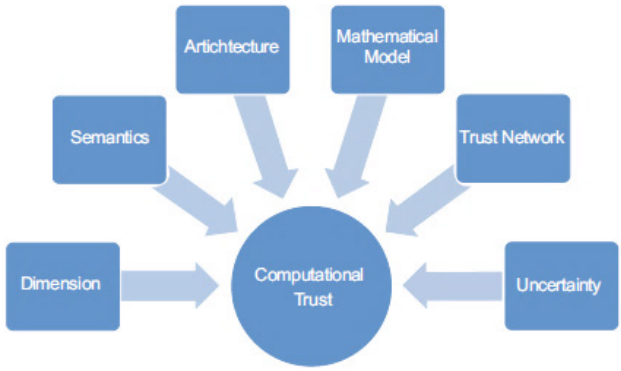


Figure 4: The Compositional Elements of Computational Trust

The authors of [13] seem to have classified trust into three tiers, i.e., Credential Base Trust Management Systems (all service providers along with the services are reliable but those requesting for it are not necessarily trust worthy), Reputation Baes Trust Management Systems (Both the providers of trust and the services are unreliable and the requesters choose the providers based on the values of reputation to provide the desired services) and Social Network Based Trust Management Systems (social networks form a foundation for these and reputation is analyzed on the basis of societal relationships). A variety of these types are intricately described in [1].

1.2. Reputation

Reputation can be defined (an accepted definition by several researchers) as: the subjective cumulative value, as identified by the requester, with the help of the judgment by others, about the non-familiar characteristic or capability of a particular node with whom the requester has never communicated before. Reputation of a vehicle can be treated as the measure of faith which other vehicles have about the sender, based on the reliability of previously transmitted data. Basically, Reputation is an old human notion: The Romans named it “reputatio”: “reputatioest vulgaris opinioubi non estveritas” [14]. Reputation, often treated as a social control mechanism [15], involves the process of telling

the truth rather than faking ones' reputation. Since reputation is not bound to specific location, the change of region has less impact on the reputation information. That is reputable people will get their value even if they change the location. Another example is that the reputation of an individual is simply rumored.

The reputation information need not be authentic all the time as it may be biased or plagiarized by malicious recommender who wishes to destroy the recommender. The above Latin quote translates to "reputation is a vulgar opinion where there is no truth" [14]. The term reputation and recommendation are used in different context. Reputation value is used to access a stranger entity but recommendation used for known familiar entity. Reputation can be categorized as centralized and decentralized. In centralized category the reputation value is computed by a trusted third party where as in later it can be computed independently each node. Since trust depends on many factors, peers need to develop differentiated trust in multiple aspects of other nodes capability. Reputation systems if used efficiently with a positive approach, can be used to nurture humble behavior and to influence persistence to contacts in ad hoc networks. The key function of the system involves gathering, distribution, and clustering of feedbacks about the behaviors of peers and colleagues. These mechanisms seem to provide incentives for the benevolent behavior aiding people to be decisive about the people they trust. Prior experience with transaction partners can be induced into the forthcoming world in order to measure their status on trust and honesty and this is very often referred to as the "Shadow of the Future". However, they can be easily invaded unrelatable and misguided ratings. And this when done on purpose, is Recommendation-Based Attack. Note that the first Web sites to introduce reputation schemes were on-line auction sites such as eBay.com.

Classification of Reputation

In an ad hoc network reputation are viewed at two levels: local and network reputation. As reputation, it can be local or global. Local reputation, the reputation of a fellow peer based on direct conversations and relationships, and global reputation, which involves local reputation along with witnesses received from other, are the two types of reputations present. They often come in handy when it is necessary to decide whether peers should be propagated for further recommendations and to check their level of trustworthy nature. These reputations can be acquired directly or indirectly [16] (refer figure 4):

a. Direct Reputation: The witness is required to directly interrogate the peer to form a reputation about him or her. For example, if peer C had already conversed with peer D, it advises peer A regarding its intuitive feeling of trust reputation of peer D. And this kind of transfer of reputational information is called Direct Reputation [16].
b. Indirect Reputation: This involves the witness peer being interrogated by a middle peer (s) instead of direct interaction by the peer who requires the reputational feedback. For example, if peer A is unaware of the creditable nature of peer D, then he/she would infer about this to his/her fellow peer B to get a feedback in order to proceed accordingly and peer B may be required to further take this to his/her colleague peer D and then its reported back to peer A and the cycle is complete.

Generally, we identify three main types of reputation here:

- Positive Reputation
- Negative Reputation
- Neutral Reputation

Here, we define positive reputation as "Reputation of the reputation queried peer obtained from a witness peer that advises that the reputation queried peer be trusted" We define negative reputation as "Reputation of the reputation queried peer obtained from a witness peer that advises that the reputation queried peer not to be trust" We define neutral reputation as "Reputation of the reputation queried peer obtained from a witness peer that is unsure about the trustworthiness of the reputation queried peer". A newcomer or to a peer who has decided to use a brand-new identity are corresponds to zero reputation, as the case may be to avoid being admitted as someone who previously misbehaved. Non trivial is also known as characterizing good reputation. A possible approach lies in the identification of a threshold which permits to discriminate the outcome of the poll as representing a good or bad reputation. It is necessary to keep in mind that the possibility of providing or obtaining a perfectly apt and matching value is a rare chance. This is because its often characterized based on a collection of responses, if its positive opinions by a number of users, it will lead to a positive reputation which takes shape from a certain mixture of responses, producing a value higher than the threshold. The same applies to negative opinions as well.

Reputation based Trust System

An interesting and challenging research can be done/ performed in Reputation based trust management in peer-to-peer systems. The soul use of trust would be to develop trust among the colleagues, ensure safe and sound transactions to increase the satisfaction. There are four main varieties of reputation-based trust systems [17]: 1) gathering of details 2) data mapping for trust evaluation 3) broadcasting and 4) decision making. In a sensing campaign, a number of tasks are produced by the main server, one of which is then chosen by a participant. The participant, for example Pi, then would preserve his/her observations for a given task J on the basis of which he/she generates a sensing report RPi for the task. These are then sent to the server for verification. The server, on the other hand, combines the reports from a plethora of such sources and produces a detailed analysis. Some of the Properties of Reputation based Trust Systems are discussed as follows:

One can identify the main properties of reputation-based trust systems on the following basis:

- Traceability: It is suitable for participants whose past behavior is made use of for the tracking and analysis of the nature and mannerism of the person. It is a key source to predict the behavior of the person in the near future.
- Freshness: The reputational value which is being assigned to an individual participant which modulates and fluctuates the faithful and loyal qualities of the person.
- Separability: The individuals do not have any control on the updation of their reputation scores. They would have no rights to maliciously interfere or forge their reputation scores.

- Exposure: Dangerous and harmful participants should be portrayed and ejected or at the least, the hostile participants should be avoided. Unknown reputation-based trust systems are well known to acquire their previous goals and targets while masking the credentials and identity of the participants. A few of them have been mentioned in [18].
- Anonymous login: The participant must have the right to login and provide the reports in an unacknowledgeable manner or in incognito mode.
- Non-associative: The report shouldn't include the participants credentials or identification details which would reveal his/her persona. Therefore, the server wouldn't be allowed to link the report to any specific individual.
- MSR unlink ability: The server shouldn't be allowed to link Multiple Sensing Reports from the same individual.
- Anonymous demonstration: The service of giving demos and reports to the server without having to reveal their original identities to it must also be facilitated [18]. The fulfilment of these goals and preventive measures against the earlier mentioned attack revolves around the efficiency of the system. Participant should be active enough to elucidate their reputation values to the server anonymously [18]. The strength of the system determines the fulfilment of these targets and resilience to the earlier mentioned attacks.

Hence, the rest of this paper is organized as follows: Section 1 discusses the definitions of trust and reputation, trust models in VANET, parameters, and their characteristics, reputation mechanisms in Peer to Peer (P2P) networks, etc. Section 2 introduces trust and reputation model comparison in detail. Section 3 discusses about open issues in trust and reputation models. In section 4, this work argues about trust and reputation mechanisms. Finally, in section 5, we conclude with a summary of contributions. In the following sections, we do not differentiate terms "vehicle," "object," "users "and "moving object". Additionally, "Trust" and Reputation" are often used interchangeably in a network trust or reputation model.

2. Trust and Reputation Infrastructure Models

A variety of mechanisms related to trust and reputation have been proposed the longer run which ids applicable to different perspectives of life like e-commerce, peer-to-peer, etc. Trust is a feeling an individual possesses when he/she believes that the other agent will handle a given task. Trust is defined as "the belief that allows individuals to be willing to react after having taken the characteristics of the providers and the underlying Internet infrastructure into consideration". Trust should be substantially based on evidence. Trust (dimensions of trust) include confidentiality, authentication, integrity, authorization, technical security and non-repudiation. In spite of its simplicity, it captivates few of the highly targeted properties of trust (found in [19] and table 4): Trust is a two-way relation between the agent and his/ her peer.

- It is a kind of decisive statement- trust or distrust
- Trust revolves around a final aim or target which is to be achieved by the trustor after relying on the trustee.
- Trust is very often descriptive and personalized. Different trustors may make differed decisions when relating to the same goal and trustee.
- Trust is uncertain and highly cloudy.

This type of work is bound to produce modularized models combining different ways of trust and reputation.

- Trust based on conversations and meetings with people in the past,
- Trust based on the roles payed among the agents,
- Trust based on the reports sent in by the witnesses regarding the attitudinal behavior of a person, and
- Authorized reputation formed from third party references.

This type of work has been known to propose a novel and innovative peer-to-peer trust model in ad hoc network systems. In the model which we have proposed [1, 20], trust is a logical connects of emotions: trustor, trustee, and goal. The predicate turns out to be true when the trustor believes the trustee for completion of the task, else it is false. Each subject is expected to possess a set of policies on the grounds of trust. These policies reflect the trustors analyzing criteria and builds up a few traits for the trustee and the environment. However, the degree of trust varies from one person to another. The accreditation for a trust relationship is as follows:

- If T's value of a trusted relationship lies between 0 and 1, then it's trust.
- If T's value of a trusted relationship lies between -1 and 0, then it is distrust.
- If T's value is equivalent to 0, then it is not trust and not distrust.

The models follow the given steps for a systematic and efficient working [21]:

a. Gathering and sorting out information related to a specific participant with the help of others opinions and recommendations.
b. Putting together all the bits and pieces of the received notification in order to compute a matching score for each peer.
c. Choosing the highly reliable and professionalized entity in the community to carry on the task smoothly and assessing the satisfaction of the user.
d. Based on the received feedback, the final step of rewarding or punishment is put forth.

Trust ratings are updated based on the compatibility of second-hand reputation information with prior reputation ratings. For more comparison and survey regarding about trust and reputation mechanisms, refer Table 1 and table 2 (refer appendix A). Three distinctive levels have been used to classify trust and reputation models from this perspective according to what was observed in trust and reputation models [19,22]:

- Level 0: Behaviors which involve cheating and plagiarism isn't considered and it mainly focuses on a massive number of agents who provide honest opinions and ratings to condemn the response from malicious agents.
- Level 1: The model perceives the agents to hide or opinionize the details though they never lie.
- Level 2: The model includes particular mechanisms to handle liars.

Though trust and reputation differ quite a lot, they're closely related and are also used to assess a persons' trustworthy nature. The common features they share are:

- Context specific: Trust and reputation both are liable on some or the other context. For example, Mike believes in John in the form of a doctor but not as a person who can fix his car. So, with respect to the case of a doctor, John is trustworthy and with respect to a mechanic, he clearly isn't.
- Multi-faceted: With reference to the same topic, it is of prime importance to generate a varied trust in varied perspectives of the potential of a person. The same goes for reputation as well. Consider this example, a customer is likely to analyze a hotel on a number of categories like food quality, cost, service, etc. The context-specificity of trust highlights that trust depends on the situation and can be characteristic, multi-faceted, etc.
- Dynamic: Trust along with reputation may deteriorate or accelerate with experience and is likely to decay with time.

1.2 Advantages and Drawbacks of different types of Trust and Reputation Management Infrastructures

The points discussed below describe the pros and cons of different varieties of trust and reputation management frameworks:

Centralized systems:

They use a reliable federal server for handling the reputation details.

Advantages:

- Flexible and simple usage
- Simplified design structure
- Ensures efficiency as all the details received by the peers are handed over to the corresponding peer.
- Keeps data together and persistent.

Drawbacks:

- Hard to attain if all peers can't trust a single entity.
- Points out even a single failure point and blockages as a large number of peers can send queries to the same entity.
- Points out even a single attack point like DoS (denial of service) attacks, Sybil attacks [8,23], sabotage and subversion.
- Costly to acquire efficient performance and strength
- Not scalable
- Not resilient to lawsuit

1.2.1 **Completely decentralized systems:** The details pertaining to the reputation is stored at their level itself and the information is scattered throughout the network.

Advantages:

- No individual failure points
- No requirements for an all-trust subject
- Provides strength and scalability
- Resistant to lawsuit

Drawbacks

- Many overhead messages are generated to maintain and handle the data
- Flooding is a necessity to acquire the required details from peers to cumulate different trust values
- The layout fluctuates often because of the momentary behavior of the peers' dude to which data loss is often existing.
- The data can be easily tampered or modified.

1.2.2 **Partially Decentralized Systems**

Advantages:

- Efficient search hacks for data
- Easily manageable and scalable
- Resilient to lawsuit
- Require a smaller number of super nodes to handle data
- Super node can easily access data from each other
- Empowering peers from the pressure of answering unnecessary doubts as peers are contacted by the respective super nodes on request alone.
- Efficient implementation of service distinction

Drawbacks:

- Super nodes should be blindly rusted to take over the details
- A burden on the nodes

Hence this paper discusses several trust and reputation computational models that attempts to address the concerns raised in table 1-4. This section discusses about the characteristics, advantages, and disadvantages of different types of trust and reputation systems. The next section contrasts several open issues in trust and reputation mechanisms in detail.

3. Open Issues

The work concludes with trust being linked to multi-agent systems by bordering the prime concerns to be tackled so as to create a complete model for an open system:

a. **Strategic Lying:** While the problem is being dealt by a few reputation models problem (such as Schillo et al. (2000) [24]; Sen & Sajja (2002) [25]; Zacharia & Maes (2000) [26]), a majority of them don't give a deeper in sight of strategized lying. It involves belittling the agents to find a trustworthy character in a liar which is then exploited for selfish deeds.

b. **Collusion Detection:** A small count of the interactive mechanism has been trained to comply with collusion (Brandt, 2002 [27]; Sen & Sajja, 2002) [25]. Furthermore, they've failed to explain how they can learn to collude though they were successful in showing how the agents could counter good deeds over time. This is the same as passing about false information to take advantage of others. We can expect agents to schematize and collude in an unrestricted environment, but if the system is expected to be efficient and well-functioning, collusion must be prevented. Or else, the agents might as well trust people who are complete liars and would take full advantage of them. In our research work, firstly, we are keen towards approaching it in a perspective which would make the system efficient and resilient to malicious mannerisms like collusion. Simultaneously, we also penetrate into the depths of combining trust control long with attack detection to suppress sudden attacks. The next concern is to detect peers as time passes

and link their past history with them. We are also continuing developing an extensive set of adaptive metrics, using the test bed to evaluate them, and finally testing the approach with real workload data.

c. **Context:** A large number of the models fail to consider the important fact that interactions often happen in an organizational context. It is absolutely wrong to label an agent dishonest or as a liar just because an agent has performed poorly, as it may be due to certain professional environment changes. Instead, the environmental variable is to be taken into consideration for this very purpose. This calls for a better risk analysis in the environment (Yamagishi et al. 1998 [28]; Molm et al. 2000 [29]). The absence of stern rules may lead to increased risk factors, which in turn will result in the analysis of an honest agent to be considered extremely trustworthy than his actual nature. Thus, if rules hinder dishonesty, there will be no requirement to increase trust related conversations among partners (Molm et al. 2000 [29]).

d. **Expectations:** Hardly any of the models surveyed are capable of conveying their expectations to one another (consider the case of data exchange regarding the quality of goods or delivery time). In an unrestricted environment, agents are free to possess different view-points and approaches which would intimate the type and form of interaction. For example, "High quality service" can mean "exact timed delivery" for one person and it may mean "good pricing" for another one. REGRET has put forth a transcendental dimension trust rating [30, 31] but this one doesn't show the conversation between the peers to comprehend each other's expectation. Once the expectation has been analyzed, the agent can satisfy the other side of the party easily. Or else, they may feel deemed to be untrustworthy with an agent.

e. **Social Networks:** Most of the security models presume a lot of scenarios which is built on the previous interactions and conversations or which are given by the agents (Schilloet al. 2000 [24]; Sabater & Sierra, (2001, 2002) [30, 32]; Yu & Singh, 2002b [33]). The semantic of the connections should be clearly defined within the network. (for example: as collaborators, partnerships in coalitions, or members of the same organizations)

f. **Coping with Peer Abuses and Selfishness:** To secure P2P system application a thorough investigation is needed for various behavior models. New mechanisms should be proposed to deal with intrusions, free riders, black mouths, collusions, and selfishness of peers. More focus should on game theoretic studies and benchmark.

g. **Reputation System for Unstructured Peer to Peer (P2P) System:** DHT (Distributed Hash Table) overlay network provides the foundation for Power Trust [34], Eigen Trust [35], and Peer Trust [36]. But, a majority of the P2P systems formed on the internet platform are unstructured and is in high demands. The major challenge faced in this field is the speedy performance of the reputation aggregation while forbidding the use of fast searching or hashing mechanism. To cross this hurdle, we are mainly focused on a gossip-based mechanization.

h. **Explore New Killer Peer to Peer (P2P) Applications:** The analysis of modern P2P applications for structured and unstructured P2P systems are a necessity. Most of the current P2P applications don't seem to have a strong and interactive bond between the users and hence, cooperation between the users should be apprehended.

i. **Others:** There are some other important concern yet to be focused are as follows:

- Impact of network dynamics on trust: The intricate analysis and description of the impact is yet to be looked into in the case of trust dynamics. The best example is that flexibility often affects the trust propagations and other modules and the clarified relationship is yet to be studied.
- Computations of trust in cooperative and non-cooperative games: In a self-organized distributed network, nodes may result in positive or negative suggestions in a straightforward or malevolent manner. These perspectives are a subdivision to a variety of scenarios in the complicated systems combined with game theoretic interactions, of which the games may be non-cooperative but are controllable using Nash equilibrium. It is to be noted that every node indulges in the games individually or in a compound manner wherein, a group of nodes form smaller groups to play games with each other and against the remaining ones. However, this field hasn't been studied to great depths.
- Impact of heterogeneous nodes on trust: Using wireless networks prove to be highly ununiform. Here, uniformity refers to the roles played by the nodes, their capacity and refuge. When trust takes form, varied evaluations ensure that not all of the nodes or their contents are treated the same. Hence, the very same descriptive pattern cannot be applied to monitor the trust levels of all nodes and their details. However, this field requires extensive knowledge and induction of network dynamics and heterogeneity in the trust evaluation functions.
- Security paradigms to enhance trust in the network: The capacity of data delivery and security concerns complements the amount of trust a recipient is likely to put forward on the data handed over. For example, the piece of information we've received can't be trusted completely if the sender or its path seem to be malicious. Furthermore, if the authenticity legalization system isn't functioning in full swing, it would be our call whether to trust the data or not. Deep dwelling into the subject would give further clarity on these issues.
- Social and context dependent trust: In the modern era, social relationships and content-based trust have been in the limelight. However, this is an area not often touched by Mobile Ad-hoc Network (MANET) because the dependence between the networks, be it social or communications, and their application are yet to be explored by MANET. Since validation of trust is a prime area for future research, they can be aided from the social communities.

This section discusses various open issues with respect to trust and reputation systems for further research. The next section arguing about trust and reputation and also provides some views to respective models.

4. Arguing about Trust and Reputation

We have discussed about trust and reputation in above sections and finally, this work made a conclusion that trust is an essential entity which is needed to be maintained in centralized system or decentralized and distributed systems. So now we discuss or arguing all possible ways to enhance trust and reputation in people and organizations.

4.1. Arguing about Trust

As definition of Trust "it is the belief the trusting agent has in the trusted agent's willingness and capability to deliver a mutually agreed service in a given context and in a given time slot". As we can be seen, trust pervades multi-agent interactions at all levels. Instead of trying hard to apprehend a translation between each other for trust evaluation, a versatile approach would be to convince others about how and why our very own evaluation seems to be apt. Arguments related to trust are mainly focused on whether an evaluation or analysis should be accepted or not. For contemplating among the models, we need to develop a new approach of looking at trust. Current models are highly colossal: input refers to the evidence which is further processed and then the output produced is the trust report. T argue about the betterment of one trust model over that of another, every agent should have a clear-cut perspective of why his very own model is the best (Pinyol et al. 2008 [37]; Castel franchi and Falcone, 2010 [38]). We consider an agent based on multi-context systems as in Pinyol et al. work (2008) [37] and have taken up the case further in a reflective manner as in [20]. Our target is to implement this in an optimized manner, allowing agents to allow for adapting to its trust model. With this, our next step is to lay down a protocol that enables agents to convince each other of the benefits of the model, limited by a specific domain in order to acquire this alignment. Gradually, the methods can be combined to provide an agent with multiple options to analyze another agents' model. In the long run, we intend to explore the interplay of the various alignment methods.

4.2. Arguing about Reputation

As deliberated earlier, reputation is the opinionized perspective of a person or an object. Oxford dictionary defines reputation to be the beliefs that are generally held about someone or something. Abdul Rahman et al. [2] define reputation as "an expectation about an agent's behavior based on information about its past behavior". Chang et al. [3] define reputation as "an aggregation of the recommendations from all of the third-party recommendations agents and their first, second and third hand opinions as well as the trustworthiness of the recommendation agent in giving correct recommendations to the trusting agent about the quality of the trusted agent". As discussed, we find out that, Trust is must needed element in human being to make love or better relationship with other human being, i.e., nothing can be operate without it (operable otherwise). Once you become trusted to someone, then you can be good for that user (in future). It is difficult to gain/maintain in case of product/organization but easy to lose. Actually, once if we lose trust among our family, it can be regained but it takes time. But if we lose our trust among any company/ corporation, it is hard to build. Once it is gone, it is difficult to recover/ build. Hence, this section provides and argues with trust and reputation models. Now next section concludes this work in brief.

5. Conclusions

Due to enhancement in Peer-to-Peer (P2P) ad hoc networking technology, Trust Mechanism has emerged to be the talk of the town. In a peer- to-peer system, trust worthiness and reputation identification play a vital role. The importance of these factors is increasing in peer-to-peer communication, since it can protect unreliable, malicious peer. Here we talked about these mechanisms in different context. And in each context different methods were used. That is there is no unique solution suitable for all scenarios. All type of constraints and input information has to be taken into account while implementing a new system. We find throughout this work, there is no single mechanism which suits for all scenarios. We have tried to showcase the different methods to identify the apt computation model to analyze the worth of the existing T&R systems across VANETs domains in this paper. Although there has been a significant number of works in T&R systems (refer table 1 and 2, in appendix A). As conclusion, a fair amount of work has been done in the area of computing reputation-based trust ratings but not sufficient according to human being's life. We need to focus in those issues and want to provide reliable experiences to human beings that no one has provided before.

Acknowledgement

This research work is funded by the Anumit Academy's Research and Innovation Network (AARIN), India. The authors would like to thank AARIN, India, a research network for supporting the project through its financial assistance.

Conflicts of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] Amit Kumar Tyagi, N.Sreenath, "Never Trust Anyone: Trust-Privacy Trade-Offs in Vehicular Ad hoc Network", British Journal of Mathematics & Computer Science (BJMCS), Vol. 19, no. 6, 2016.

[2] A. Abdul-Rahman, S. Hailes, "Supporting Trust in Virtual Communities," Proc. 33rd Hawaii International Conference on System Sciences, 2000.

[3] Chang, E.,Dillon, T.,Hussain, F.K.,"Trust and Reputation for Service-Oriented Environments". Wiley, 2006.

[4] D. H. McKnight and N. L. Chervany, "What is Trust? A Conceptual Analysis and an Interdisciplinary Model", in Proceedings of AMCIS, 2000.

[5] D. McKnight and N. L. Chervany, "The Meanings of Trust." MISRC 96-04, University of Minnesota, Management Informations Systems Research Center, 1996.

[6] T. Grandison and M. Sloman,"A Survey of Trust in Internet Applications". IEEE Communications Surveys and Tutorials, 3, 2000.

[7] Zhang, J. "A survey on trust management for VANETs", In Proceedings of the 25thIEEE International Conference on Advanced Information

Networking and Applications, AINA 2011, Biopolis, Singapore, pp. 105–112, march 2011.
[8] Amit Kumar Tyagi, N. Sreenath, "Preserving Location Privacy in Location Based Services against Sybil Attacks", International Journal of Security and Its Applications (IJSIA), Vol.9, No.12, pp.189-210, December, 2015.
[9] M.K. Reiter and S.G. Stubblebine, "Toward acceptable metrics of authentication". In Proceedings of the 1997 IEEE Symposium on Research in Security and Privacy, Oakland, CA, 1997.
[10] Mauro Stocco, Thomas Engel, Uwe Roth, "Trust Arrays: Allowing P2P nodes to "personally" evaluate trustworthiness of potential partners", Advances in Intelligent Systems - Theory and Applications, In cooperation with the IEEE Computer Society, Luxembourg, November, pp. 15-18, 2004.
[11] F. Bao and I.R. Chen, Dynamic Trust Management for the Internet of Things Applications, International Workshop on Self-Aware Internet of Things, San Jose, USA, September, 2012.
[12] Yu, Bin Yu Bin Singh, M.P. Sycara, K."Developing Trust in Large-Scale Peer-to-Peer Systems",IEEE First Symposium on Multi-Agent Security and Survivability, 2004.
[13] Suryanarayana, G. Taylor, R.N. "A Survey of Trust Management and Resource Discovery Technologies in Peer-to-Peer Applications". Tech. rep. ISR, 2004.
[14] Bouvier M. "Maxims of Law", Law Dictionary, 1856.
[15] Kuwabara K. "Reputation: Signals or incentives?", In The annual meeting of the American sociological association, 2003.
[16] Patel, Nirav J Jhaveri, and Rutvij H, "Trust based approaches for secure routing in VANET: A Survey", International Conference on Advanced Computing Technologies and Applications, 2015.
[17] K. J. Hoffman, D. Zage, C. Nita-Rotaru, "A survey of attack and defense techniques for reputation systems", ACM Computing Surveys (CSUR), Vol. 42, no. 1.
[18] X. O. Wang, W. Cheng, P. Mohapatra, T. F. Abdelzaher, "Artsense: Anonymous reputation and trust in participatory sensing", in: INFO- COM, IEEE, pp. 2517–2525, 2013.
[19] Amit Kumar Tyagi, N. Sreenath, "Providing together Security, Location Privacy and Trust for moving objects", International Journal of Hybrid Information Technology (IJHIT) Vol.2, No.3, pp. 221-240, March, 2016.
[20] Amit Kumar Tyagi, N.Sreenath, "Providing Trust Enabled Services in Vehicular Cloud Computing (extended version)", in proceeding of ACM/International Conference on Informatics and Analytics (ICIA), pp. 25-26 August, Pondicherry, India, 2016. DOI: http://dx.doi.org/10.1145/2980258.2980263.
[21] Marti, S. and Garcia-Molina, H. "Taxonomy of trust: categorizing P2P reputation systems", Computer Networks, Vol. 50 No. 4, pp. 472-484, 2006.
[22] Sabater, Jordi Sierra, Carles, "Review on computational trust and reputation models", Artificial Intelligence Review, Springer, Vol. 24, Issue 1, pp 33-60,September 2005.
[23] Amit Kumar Tyagi, N. Sreenath, "Providing Safe, Secure and Trusted Communication among Vehicular Ad hoc Networks' Users: A Vision Paper", International Journal of Information Technology and Electrical Engineering, Vol. 5, Issue 1 February 2016.
[24] Schillo, M. Funk, P.&Rovatsos, M."Using trust for detecting deceptive agents in artificial societies", Applied Artificial Intelligence, Special Issue on Trust, Deception, and Fraud in Agent Societies, Vol. 14, no. 8, pp. 825-848, 2000.
[25] Sen, S.& Sajja, N."Robustness of reputation-based trust: Boolean case", In Castelfranchi, C.& Johnson, L.(eds.), Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems, Vol.1, pp. 288-293, 2002.
[26] Zacharia, G.&Maes, P. "Trust through reputation mechanisms", Applied Artificial Intelligence 14, pp.881–907, 2000.
[27] Brandt, F. "A verifiable bidder-resolved auction protocol", In Workshop on Deception, Fraud and Trust in Agent Societies, AAMAS, Bologna, Italy, pp.18–25, 2002.
[28] Yamagishi, T. Cook, K. & Watabe, M."Uncertainty, trust, and commitment formation in the United States and Japan", American Journal of Sociology, 104, pp. 165-194, 1998.
[29] Molm, L. D. Takahashi, N.& Peterson, G."Risk and trust in social exchange: an experimental test of a classical proposition", American Journal of Sociology 105, pp. 1396-1427, 2000.
[30] Sabater J. and Sierra C. "REGRET: a reputation model for gregarious societies". In 4th Workshop on Deception, Fraud and Trust in Peer Societies, 2001.
[31] Sarvapali D. Ramchurnet. al. "Trust in multi-agent systems", The Knowledge Engineering Review, Vol. 19, no. 1, pp. 1-25, Cambridge University Press,2004.
[32] Sabater, J.& Sierra, C."REGRET: a reputation model for gregarious societies", In Castelfranchi, C.&Johnson, L.(eds.), Proceedings of the 1stnternational Joint Conference on Autonomous Agents and Multi-Agent Systems, pp. 475-482 (Extended work of 2001), 2002.
[33] Yu, B.& Singh,M. P."An evidential model of reputation management", In Castelfranchi, C.& Johnson, L.(eds.), Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems, Vol. 1, pp. 295-300, 2002b.
[34] Zhou, R. and Hwang, K. "PowerTrust: a robust and scalable reputation system for trusted peer-to-peer computing", IEEE Transactions on Parallel and Distributed Systems, Vol. 18 No. 4, pp. 460-473, 2007.
[35] Kamvar, S. Schlosser, M. and Garcia-Molina, H. "The Eigen-Trust algorithm for reputation management in P2P networks", WWW03: Proceedings of the 12thinternational conference on World Wide Web, pp. 640-651, 2003.
[36] Xiong, L. and Liu, L. "PeerTrust: supporting reputation-based trust in peer-to-peer communities", IEEE Transactions on Knowledge and Data Engineering, Vol. 16 No. 7, pp. 843-857, 2004.
[37] Pinyol et al. I. Pinyol, J. Sabater-Mir, and P. Dellunde, "Cognitive social evaluations for multi-context bdi agents". In CCIA'08, 2008.
[38] Castelfranchi and Falcone, C. Castelfranchi and R. Falcone,"Trust Theory: A Socio-cognitive and Computational Model", Wiley, 2010.
[39] Go´mez, F. and Martı´nez, G. "Providing trust in wireless sensor networks using a bio-inspired technique", Telecommunications Systems Journal, Vol. 46, no. 2, 2010.
[40] Go´mez, F. and Martı´nez, G. "TRMSim-WSN, trust and reputation models simulator for wireless sensor networks", paper presented at IEEE ICC09: International Conference on Communications, 2009a.
[41] Anupam Das and Mohammad Mahfuzul Islam, "SecuredTrust: A Dynamic Trust Computation Model for Secured Communication in Multiagent Systems", IEEE Transactions on Dependable And Secure Computing, Vol. 9, no. 2, March/April 2012.
[42] Yonghong Wang, Munindar P. Singh, "Evidence-Based Trust: A Mathematical Model Geared for Multiagent Systems," ACM Transactions on Autonomous and Adaptive Systems, Vol. 5, No. 3, pp. 1-25, September 2010.
[43] Justin Manweiler et al., "SMILE: Encounter-Based Trust for Mobile Social Services," CCS'09, November 9-13, Chicago, Illinois, USA, 2009.
[44] Heba A. Kurdi, "HonestPeer: An enhanced Eigen-Trust algorithm for reputation management in P2P systems," Journal of King Saud University-Computer and Information Sciences, 27, pp. 315-322, 2015.
[45] Marsh, S.P. "Formalising trust as a computational concept", PhD thesis, Department of Computing Science and Mathematics, University of Stirling, Stirling, 1994.
[46] Go´mez, F. and Martı´nez, G. "State of the art in trust and reputation models in P2P networks", in Shen, X. Yu, H. Buford, J. and Akon, M. (Eds), Handbook of Peer-to-Peer Networking, Springer, New York, NY, pp. 761-784, 2009b.
[47] Mui, L. "Computational models of trust and reputation: agents, evolutionary games, and social networks", PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 2002.
[48] Sun, Y.L. and Yang, Y. "Trust establishment in distributed networks: analysis and modelling", IEEE ICC07: International Conference on Communications, pp. 1266-1273, 2007.
[49] Lam, S.K. and Riedl, J. "Shilling recommender systems for fun and profit", WWW '04: Proceedings of the 13thInternational Conference on World Wide Web, 2004.
[50] Go´mez, F. and Martı´nez, G. "TRMSim-WSN, trust and reputation models simulator for wireless sensor networks", paper presented at IEEE ICC09: International Conference on Communications, 2009a.
[51] Moloney, S. "Simulation of a distributed recommendation system for pervasive networks", SAC05: Symposium on Applied Computing, pp. 1577-1581, 2005.
[52] Chen, X. Zhao, K. and Chu, X. "SepRep: a novel reputation evaluation model in peer-to-peer networks", Proceedings of the 5thInternational Conference on Autonomic and Trusted Computing, pp. 86-99, 2008.
[53] Wang, Y. Tao, Y. Yu, P. Xu, F. and Lu, J. "A trust evolution model for P2P networks", Proceedings of the 4thInternational Conference on Autonomic and Trusted Computing, pp. 216-225, 2007.
[54] Gui, C. Wu, Q. and Wang, H. "Towards trustworthy resource selection: a fuzzy reputation aggregation approach", Proceedings of the 4thInternational Conference on Autonomic and Trusted Computing, pp. 239-248, 2007.
[55] Gui, C. Wu, Q. Wang, H. and Qiang, J. "Towards trustworthiness establishment: a D-S evidence theory based scorer reliability tuned method

for dishonest feedback filtering", Proceedings of the 5thInternational Conference on Autonomic and Trusted Computing, pp. 444-454, 2008.
[56] Chen, R. Zhao, X. Tang, L. Hu, J. and Chen, Z. "CuboidTrust: a global reputation-based trust model in peer-to-peer networks", Proceedings of the 4th International Conference on Autonomic and Trusted Computing, pp. 203-215, 2007.
[57] Azzedin, F.Ridha, A. and Rizvi, A. "Fuzzy trust for peer-to-peer based systems", Proceedings of World Academy of Science, Engineering and Technology, Vol. 21, pp. 123-7, 2007.
[58] Hao Hu, Rongxing Lu, et al. "REPLACE: A Reliable Trust-based Platoon Service Recommendation Scheme in VANET", IEEE Transactions on Vehicular Technology, 2016.
[59] Serna, Jetzabel, Jesus Luna, and Manel Medina, "GeolocationBased Trust for Vanet's Privacy", In Information Assurance and Security, ISIAS'08, Fourth International Conference on, pp. 287-290. IEEE, 2008.
[60] LinkeGuo, Member, Chi Zhang et al., "A Trust-Based Privacy-Preserving Friend Recommendation Scheme for Online Social Networks", IEEE Transactions on Dependable And Secure Computing, Vol. 12, no. 4, July/August 2015.
[61] Fe´ lix, Go´mez Ma´rmol et al., TRIP, a trust and reputation infrastructure-based proposal for vehicular ad hoc networks", Journal of Network and Computer Applications 35, pp. 934–941, 2012.
[62] Huynh et al. et al., "An integrated trust and reputation model for open multi-agent systems". Journal of Autonomous Agents and Multi-Agent Systems, Vol. 13, no. 2, pp. 119-154, September 2006.
[63] B. Yu and M. P. Singh, "Detecting deception in reputation management", in Proceedings of the 2ndInternational Joint Conference on Autonomous Agents & Multiagent Systems, Melbourne, Australia, ACM, pp. 73-80, 2003.
[64] W. T. L. Teacy, J. Patel, N. R. Jennings, and M. Luck, "TRAVOS: Trust and reputation in the context of Inaccurate information sources, Autonomous Agents and Multi-Agent Systems", Vol. 12, no. 2, pp. 183-198, 2006.
[65] A. Jøsang, and R. Ismail, "The beta reputation system", in Proceedings of the 15th Bled Conference on electronic Commerce, Bled, Slovenia, 2002.
[66] Hong, Xiaoyan, Dijiang Huang, Mario Gerla, and Zhen Cao, "SAT: situation-aware trust architecture for vehicular networks", In Proceedings of the 3rdinternational workshop on Mobility in the evolving internet architecture, pp. 31-36, ACM, 2008.
[67] Wang, Zhou, and ChunxiaoChigan, "Countermeasure uncooperative behaviors with dynamic trust-token in VANETs", In Communications, ICC'07, IEEE International Conference on, pp. 3959-3964, IEEE, 2007.
[68] Biswas, Subir, JelenaMisic, and Vojislav Misic, "ID-based safety message authentication for security and trust in vehicular networks", In Distributed Computing Systems Workshops, 31st International Conference on, pp. 323-331. IEEE, 2011.
[69] Wei, Yu-Chih, and Yi-Ming Chen, "An efficient trust management system for balancing the safety and location privacy in VANETs", In Trust, Security and Privacy in Computing and Communications (TrustCom), IEEE 11thInternational Conference on, pp. 393-400, IEEE, 2012.
[70] Amit Kumar Tyagi, N. Sreenath, "A Comparative Study on Privacy Preserving Techniques for Location Based Services", British Journal of Mathematics & Computer Science (BJMCS), Vol. 10, no. 4, pp. 1-25, Article no.BJMCS.16995, ISSN: 2231-0851, July 2015.
[71] Dhurandher, Sanjay K. et al., "Securing vehicular networks: a reputation and plausibility checks-based approach", In GLOBECOM Workshops (GC Workshops), IEEE, pp. 1550-1554, 2010.
[72] Gazdar, Tahani et al."Secure clustering scheme based keys management in VANETs", In Vehicular Technology Conference (VTC Spring), IEEE 73rd, pp- 1-5. IEEE, 2011.
[73] Ding, Qing, Xi Li, Ming Jiang, and XueHai Zhou, "Reputation-based trust model in vehicular ad hoc networks", In Wireless Communications and Signal Processing (WCSP), International Conference on, pp- 1-6, IEEE, 2010.
[74] Wang, Jian, Yanheng Liu, Xiaomin Liu, and Jing Zhang, "A trust propagation scheme in VANETs", In Intelligent Vehicles Symposium, IEEE, pp. 1067-1071, 2009.
[75] Serna, Jetzabel, Jesus Luna, and Manel Medina, "GeolocationBased Trust for Vanet's Privacy", In Information Assurance and Security, ISIAS'08, Fourth International Conference on, pp. 287-290. IEEE, 2008.
[76] Gazdar, Tahani, AbderrahimBenslimane, et al., A trust-based architecture for managing certificates in vehicular ad hoc networks", In Communications and Information Technology (ICCIT), International Conference on, pp.180-185,IEEE, 2012.
[77] Chen, Yi-Ming, and Yu-Chih Wei, "A beacon-based trust management system for enhancing user centric location privacy in VANETs", Communications and Networks Journal, Vol. 15, no. 2 pp- 153-163, 2013.
[78] Chen, Chen, Jie Zhang, Robin Cohen, and Pin-Han Ho, "Secure and efficient trust opinion aggregation for vehicular adhoc networks", 72nd Conference Fall in Vehicular Technology (VTC 2010-Fall), pp. 1-5, IEEE, 2010.
[79] Sahoo, RashmiRanjan, Rameswar Panda, et al., "A trust based clustering with Ant Colony Routing in VANET", In Computing Communication & Networking Technologies (ICCCNT) Third International Conference on, pp. 1-8, IEEE, 2012.
[80] M. Raya, P. Papadimitratos, V. Gligor, and J. Hubaux, "On data-centric trust establishment in ephemeral ad hoc networks," in Proceedings ofIEEE Infocom, 2008.
[81] F. Dotzer, L. Fischer, and P. Magiera, "Vars: A vehicle ad-hoc network reputation system," in Proceedings of the IEEE International Symposiumon a World of Wireless, Mobile and Multimedia Networks, 2005.
[82] P. Golle, D. Greene, and J. Staddon, "Detecting and correcting malicious data in VANETs," in Proceedings of VANET, 2004.
[83] U. F. Minhas, J. Zhang, T. Tran, and R. Cohen, "Towards expanded trust management for agents in vehicular ad-hoc networks," InternationalJournal of Computational Intelligence Theory and Practice (IJCITP), Vol. 5, no. 1, 2010.
[84] C. Chen, J. Zhang, R. Cohen, and P.-H. Ho, "A trust-based message propagation and evaluation framework in VANETs," in Proceedings of theInt. Conf. on Information Technology Convergence and Services, 2010.
[85] M. Gerlach, "Trust for vehicular applications," in Proceedings of the International Symposium on Autonomous Decentralized Systems, 2007.
[86] Patwardhan,A.; Joshi,A.; Finin, T.; Yesha, Y. "Adata intensive reputation management scheme for vehicular ad hoc networks", In Proceedings of the 3rdAnnual International ICST Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, MOBIQUITOUS 2006, San Jose, CA,USA, pp. 1–8, 17-21 July 2006.
[87] F. Bao and I.R. Chen, "Trust Management for the Internet of Things and Its Application to Service Composition", IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, San Francisco, USA, pp. 1-6, June 2012.
[88] D. Chen, G. Chang, D. Sun, J. Li, J. Jia, and X. Wang "TRM-IoT: a trust management model based on fuzzy reputation for Internet of Things." Computer Science and Information Systems, Vol. 8, no. 4, pp. 1207-1228, 2011.
[89] F. Bao, I.R. Chen and J. Guo, "Scalable, Adaptive and Survivable Trust Management for Community of Interest Based Internet of Things Systems", 11th International Symposium on Autonomous Decentralized System, Mexico City, Mexico, March 2013.
[90] Z. Chen, R. Ling, C.M. Huang and X. Zhu, "A scheme of access service recommendation for the Social Internet of Things," International Journal of Communication Systems, February 2015.
[91] M. Nitti, R. Girau, and L. Atzori, "Trustworthiness Management in the Social Internet of Things," IEEE Transactions on Knowledge and Data Management, Vol. 26, no. 5, pp. 1253-1266, 2014.
[92] Y.B. Saied, A. Olivereau, D. Zeghlache, and M. Laurent, "Trust management system design for the Internet of Things: A context-aware and multi-service approach," Computers and Security, 2014.
[93] I.R. Chen, J. Guo, F. Bao, "Trust Management for SOA-based IoT and Its Application to Service Composition", IEEE Transactions on Service Computing, 2015.

Appendix A

Table 1 Trust and Reputation Model Comparison (based on Proposed Model in [20])

Trust and /or Reputation model steps		**Selected Trust/or Reputation models**								
		BTRM-WSN [39]	**TRM Sim WSN [40]**	**EIGEN TRUST [35]**	**PEER TRUST [36]**	**POWER-TRUST [34]**	**SECURED TRUST [41]**	**EVIDENCE- TRUST [42]**	**SMILE-TRUST [43]**	**HONEST PEER [44]**
	Gather information	Ants leave pheromone trace during their expedition.		$C=\{c_{ij}\}$, is the matrix form built by a node.	Every client gathers responses from every other client to analyse their credibility Cr(v).	Each server i collects r_{ji} and v_j from each interacted client j	$F\ Cret\ n\ (p, q) = (1-ln(Simt\ n(p,q))/\ ln\ \theta$, *if* $Simt\ n(p, q) > \theta$ $=0$, *else*	$F_{p^1}{}^{p^2}f(p)dp.$	Number of recipients $= n/2^l = z \cdot k/(p \cdot (1-c)\ r \cdot d)$	
	Score and ranking	Every path is scored in the following manner: $Q(S_k)$.		Vector $\vec{t}_i(k)$ is computed for each node i	Every client Audits T(u) for each server-u.	Each server i, computes v_i		$c(r, s) = 0.5*f_0{}^l\ (x^r(1-x)^s\ /\ f_0{}^l\ (x^r(1-x)^s\ dx$	Overhead reception rate *(score)* $= z \cdot k/p \cdot (1-c) \cdot d) - r$	Score=*#(valid files received by good peers)/#(transactions attempted by good peers)*
	Entity selection	The highest quality path makes it to the top of the chart.		Probabilistic with probability $\frac{t_i}{\sum_j t_j}$	The server which satisfies $\max_u\{T(u)\}$ is chosen.	Server k with $max_k\{V_k\}$ is picked.				
	Transaction	The client computes her responses with the services obtained.		The client analyses her remarks with the reserved space.	The client analyses her remarks with the reserved space.	The client assesses her satisfaction with the received services.				
	Punish and reward	Pheromone evaporation		Not applicable	Not applicable	Not applicable	$DT^t{}_n\ (p, q) = Sat_n{}^t\ (p, q)$			

Note- Empty Space- Not Sure

Table 2 Review of Existing Trust and Reputation Model Literature

Category	Authors	Description
Trust and reputation management theory	Marsh (1994) [45]	The very first work related to this subject. It's a thesis bringing out bases of trust and reputation.
	Go´mez and Martı´nez (2009b) [46]	This work is mainly related to the P2P networks and doesn't implement any model or provide any surveyed comparison.
	Mui (2002) [47]	An age old work in the same field related to multi-agent systems.
	Sun and Yang (2007) [48]	An amazing piece of work formalizing the perspectives of trust and reputation management.
	Lam and Riedl (2004) [49]	The work takes us through the vulnerabilities based on the recommendation systems.
	Go´mez and Martı´nez (2009a) [50]	This paper portrays the security threats in the trust and reputation systems along with ways to overcome them.
	Marti and Garcia- Molina (2006) [21]	This work defines the general steps to be kept in mind in P2P networks.
Trust and reputation models simulation frameworks	Moloney (2005) [51]	A paper which portrays the simulations of a simple recommendation system for persistent networks.
	Go´mez and Martı´nez (2009a) [50]	It describes about the open source models like TRMSim-WSN.
Trust and/or reputation models	Chen et al. (2008) [52]	It talks about the reputation models for a P2P network.
	Wang et al. (2007) [53]	This paper puts forth a trusted model for P2P networks considering the direct experiences.

	Kamvar et al. (2003) [35]	Aiming to develop P2P network, this paper has proposed EigenTrust which is the highly used model.
	Go´mez and Martı´nez (2010) [39]	BTRM-WSN is a part of the paper, which is a rare piece making use of ant colony system to handle trust and reputation in wireless networks.
	Gui et al. (2007) [54]	This work proposes an aggregate of the reputation by using fuzzy sets and logics.
	Gui et al. (2008) [55]	It shows methods for filtering of dishonest and malicious feedbacks using Dempster-Shafer theory.
	Xiong and Liu (2004) [36]	One of the highly cited papers, it's based on peer trust in P2P networks in the field of e-commerce systems and gives two ways in the form of credibility measures. It also satisfies community context factors in trust measurement, employing a camouflaging design for the peer location. It supports a single-rating system.
	Chen et al. (2007) [56]	CuboidTrust is portrayed in this paper which is handy for P2P networks. It build strong bonds and ties between contribution, trustworthiness and qulity of resources, applying power to analyse the global trust value of each peer.
	Zhou and Hwang (2007) [34]	PowerTrust is presented in this paper and is considered to be an enhancement of EigenTrust receiving international attention.
	Azzedin et al. (2007) [57]	This paper uses uncertain logics and reasonings to handle ambiguous details of trust in P2P networks.
	Abdul-Rahman and Hailes (2000) [2]	It is one of the first few works and describes the general properties of trust and reputation systems.
	TassosDimitriou, et al., (2007) [58]	SuperTrust, which is not a centralised approach, is based on K-redundant peer networks and guarantees a complementary support to the proposed models for developing trust.
	Hao Hu,RongxingLu , et al., (2016) [58]	REPLACE, a reputation system, has been designed for the team headed vehicles by gathering and prototyping the users' feedback. This is a scure and sound method against frequent attacks, belittling, etc.
	Anupam Das, Mohammad Mahfuzul Islam (2012) [41]	SecuredTrust, which identifies the malicious alterations in behavior along with striking a balance of work among the service providers is introduced here. It keeps in mind a variety of factors to determine the trust of an individual like similarities, feedback and responses, previous trust issues, etc. and hence remains more effective amidst a large number of agents.
	Jetzabel Serna, Jesus Luna and Manel Medina (2009) [59]	Geolocation-Based Trust passes on the details related to a vehicles' location.
	LinkeGuo, Member, Chi Zhang et al., (2015) [60]	This paper, introduces a new field, i.e., trust with privacy. It's a trust-based privacy preserving friend recommendation scheme for OSN's. The OSN users use their traits to identify matching peers so as to develop a social relationship with the unknown through a trust-chain.
	Fe´ lix Go´mezMa´rmol et al., (2012) [61]	TRIP, a quick, scalable model, is used for resolving on whether to accept a traffic input dropped in by another vehicle or not by analysing the loyal nature of the issuer of the message.
	Xinlei, Wei Cheng, et. al (2014), [33]	ARTSense is a skeletonised system to overcome the issue of trust without identity in mobile sensing. There's no need for the presence of a trusted third party and both positive and negative reputational messages can be applied. a framework to solve the problem of "trust without identity" in mobile sensing. In this, no require of trusted third party and both positive and negative reputation updates can be enforced.
	(T.D Huynh, N.R. Jennings, N.R.Shadbolt, 2006)[62]	FIRE focuses on the multi-agent system with the help of four information reservoirs and deals with the problems of freshers and eradicates the unnecessary details along with distinguishing between dishonest and mistaken agents. It also provides reliability measures by using an overrated rating system approach to complement MAS.
	(JordiSabater and Carles Sierra, 2002) [19]	REGRET is meant for sophisticated e-commerce platforms and is into the development of sociogram and modelized social relationships. It complements system reputation and provides a huge dimension to bring together different perspectives of behavior in terms of reputation. It evaluates honesty through unambiguous rules and provides a measure of reliability to employ an advanced rating system.
	(B. Yu, M.P. Singh,2003) [63]	This work is highly useful for the MAS (multi-agent system) as it provided novel trust and reputed network and is capable of identifying three models of description. It helps in distinction of agents who have good or no reputation at all with the Dempster-Shafter theory and it highly suffices dynamism in MAS.
	(W. T. L. Teacy, J. Patel, et al,2006) [64]	TRAVOS has been created for a huge scale open system, providing two sources for information. It takes good advantage of the probabilistic approach to obtain the credibility of the witnesses and provides confidence and reliability measures for apt conversations between the information sources. It makes use of a single rating system.
	(A.Jøsang,, 2002) [65]	Beta Reputation System (BRS), which is highly advised for a dynamic surrounding, supports binomial rating systems and comprehends the bootstrapping issues by keeping in mind the quality of the society in the marketplace, provide repeated filtration algorithms which is capable of unveiling malicious intentions if majority of the participants act loyally, exploit the longevity factor to discount the ratings with time, allow participants to fluctuate their mannerisms to increase their own profits.

Table 3 Summary of Trust Establishment Techniques in VANETs

Topic Name	Description	Mechanism / Algorithm	Methodology	QoS / Performance
SAT: Situation-Aware Trust architecture for vehicular Networks [66]	It's the middle layer forming a strong relation among the nodes consisting of SAT and STL.	SAT architecture (Situation-Awareness Trust)	Developing trust which revolves around cryptographic methods like data integrity and authentication.	Managing policy, trust improvement, Social network
Analyzing the credibility of the guards in VANETs [6] for safety.	Watchdogs keep an eye on the neighboring nodes and observe the behavioral values.	Watchdog algorithm with intrusion detection	Neighboring nodes carry on the packets forward and guard the nodes.	Detection potential, false negative and false positive
Counter the Unacceptable behaviors with dynamic trust-token in VANETs [67]	DTT puts forth a carefully calculated trust at real time functioning of the node spontaneously	Dynamic Trust-Token (DTT)	A variety of Symmetric and asymmetric cryptographic methods for veracity and watchdog is used for creating trust tokens.	Protect packet integrity, potential degradation
ID-based safety message authorization for safety in Vehicular Networks [68]	ECDSA is useful for the RSU unit authorization and for verification purpose	ID-based proxy signature and ECDSA	For message authentication and trust control, certificates are made less public.	Message transfer with authentication throw trusted RSU
A structured and systematic trust management system for maintain safety and location privacy in VANETs [69, 70]	RSU makes an instantaneous decision to identify the trust worthiness of the message being sent by the vehicle and it also prevents internal attacks.	Road-side unit (RSU) and Beacon-Based Trust System (RABTM)	Event-based trust systems are used for developing trust and beacon message and event message to govern the value of trust for that event.	Safety and location privacy of vehicles
TRIP, a trust and reputation Infrastructure based proposal for vehicular ad hoc networks [61]	Predict the destructive and harmful nodes which are likely to broadcast unwanted data. This is achieved by developing an update in the central database in a recurring fashion.	Trust and Reputation Infrastructure based Proposal (TRIP)	Anomalous sets divide trust and categorized servity of past trust	Recognizing spoilt and destructive nodes which spread false details.
Securing vehicular networks: a reputation and possibility checks-based approach [71]	Opinion generation, opinion piggybacking and provision of node reputation which can be trusted or not. False message creation identifies with PVM.	Vehicular Security produces Reputation and Plausibility check (VSRP), VARs algorithm	VARs algorithm showcases indirect and direct trust, reputation-based algorithm	Event Modification, false event production, data aggregation and data Dropping
Secure clustering scheme-based keys management in VANETs [72]	VDDZ popularises filter certificate request which is provided by CA in the group and protects direct interaction and destroys attacks.	VANET Dynamic Demilitarized Zone (VDDZ)	Divide cluster head (CH) of neighbor node and Registration authority (RA) provides the confidence to neighbors	Prevent destructive and anomalous vehicles within cluster
Reputation-based trust model in vehicular ad hoc Networks [73]	All vehicles come across same traffic issues and differentiate the roles occurred in the event.	Event based reputation algorithm	Random way point scheme to adopt for identify bogus information	Filter bogus and false warning message, enhance trust
A trust propagation scheme in VANETs [74]	Creating a serene relation between the past and current trust status, analyzing trust on the grounds of attributes and processing the similarity between the two nodes.	Novel scheme for enhancing trust management	Attributes comparison with trust value	Enhancing trust propagation, reliable packet delivery
Geolocation-Based Trust for VANET's Privacy [75]	Strict access management provides trust verification among nodes. New techniques result in valid trust geographical areas.	Geolocation based establishment	Pseudonyms used for privacy and MAC trusted location	Privacy mechanism
A trust-based architecture for managing certificates in vehicular ad hoc networks [76]	Certificate authority (CA) provides authorization legitimately to vehicles and uncertain distinguee honest node and clusters broadcasts the trust values to its neighbors.	Fuzzy algorithm, Certification Authority (CA)	Fuzzy based solution and certificate authority (CA) and PKI scheme	CA within cluster only Co-operation with vehicles and legitimate broadcast data
A beacon-based trust management system for advancing user centered location privacy in VANETs [77]	BTM generates and verifies the position of the vehicles and its direction. Message transmission is used for cryptography and pseudo identity scheme	Beacon-based Trust Management (BTM), Dempster Shafer evidence	Beacon establishes trust relationships, FSP and RSP along with location privacy enhancement scheme	internal attacks, bogus message and privacy enhancement

Secure and efficient trust opinion aggregation for vehicular ad hoc networks [78]	A mixture of a lot of message signs into a single one is forwarded to the vehicle	Identity based Aggregate algorithm	Trust opinion aggregate scheme	Space efficiency and time complexity
A trust-based clustering with Ant Colony Routing in VANET [79]	CH roots the networks in a group and calculates the indirect trust on the nodes	Trust dependent Ant Colony Routing (TACR), Mobility aware Ant Colony Optimization Routing (MARDYMO)	MAR-DYMO used for routing overhead in the network. CH calculates the indirect trust value, MAR-DYMO, for optimized routing technique	Scalability, real time updated position and trust value of vehicles

Table 4 Properties of the Existing (Some) Trust Model for VANETs

Approaches	Raya et al. 2008 [80]	Dotzer et al. 2005 [81]	Golle et al. 2004 [82]	Minhas et al. 2010b [83]	Chen el al. 2010 b [84]	Gerlach 2007 [85]	Pawardhan et al. 2006 [86]	Fe´ lix Go´mezMa´rmol et al. 2012 [61]	Hao Hu, et al. 2016 [58]
Decentralized	Y	Y	Y	Y	Y		Y	Y	Y
Sparsity			Y	Y	Y	Y	Y		
Dynamics	Y	Y		Y	Y	Y	Y		
Scalability				Y	Y				Y
Confidence	Y			Y	Y	Y			
Security	Y		Y	Y	Y	Y	Y	Y	
Privacy		Y	Y	Y		Y			
Robustness			Y						

Note:Y-Yes, Empty Space - Not Sure

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 403-407 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Classification System for the Interpretation of the Braille Alphabet through Image Processing

Brian Meneses-Claudio*, Witman Alvarado-Diaz, Avid Roman-Gonzalez

Universidad de Ciencias y Humanidades, Image Processing Research Laboratory (INTI-Lab), Perú

ARTICLE INFO

Article history:
Received: 15 October, 2019
Accepted: 28 January, 2020
Online: 20 February, 2020

Keywords:
Braille Language
Image Processing
Point Detection
Visual Disability
Braille Language Translation
Peruvian Vocabulary
Braille Language Specialists

ABSTRACT

The Braille language system or also known as the language for the blind people was created so that these people can understand the vocabulary words depending on the country in which they are located, this system is based on points in which the points raised or also known as the largest points depending on their location will have a meaning of a letter. At present, very few educational centers have Braille translation systems or Braille language specialists for the understanding and teaching of people with visual disabilities. According to INEI, 1.9% of the Peruvian population has visual impairment and only 20% of that population manages to finish their studies thus having a great deficit in cognitive preparation. That is why a Braille translation system was proposed, capable of capturing images in Braille language and translating them through image processing into Peruvian vocabulary, giving as a result, the understanding of the person or specialist who is teaching or training in that moment. The Braille language image has to go through different filters of which are intended to obtain the areas of the circles that compose it in order to obtain a binary matrix of it and convert it to a number that will represent the letter. The result was the translation of the Braille language into Peruvian vocabulary and also the identification of the points that have each letter.

1. Introduction

The Braille system or also known as language for the blind is used by people who suffer or have some visual impairment or blindness, they require learning the Braille system to write and read texts, books and documents, this reading and writing system is tactile. These people tend to learn through audio but not by practicing motor hand movements, therefore, they will not be able to write or read texts in Braille [1].

According to the INEI (Instituto Nacional de Estadísticas e Informática), there are 600,000 people with visual disabilities in Peru, of which 160,000 suffer from general blindness and the other 440,000 have a disease related to blindness or loss of the vision [2]. However, being a large population, many public entities still do not integrate technology, training and rehabilitation systems so that young people with these ills can integrate and access primary, secondary or higher education.

The importance of the knowledge of the Braille system will allow the integration of these systems in the curricula for young people because it is estimated according to the INEI that 23.9% of the total population of people with visual disabilities manage to complete their studies, this problem implies that many blind people cannot work despite the fact that there is currently a general law 29973, which stipulates the integration of people with disabilities in the workplace, sectoring 5% in state entities and 3% in private companies [3].

The implementation of Braille Language Translation systems can help to reduce the gap that exists in the lack of knowledge of the same language and the knowledge of the importance of this language for people with visual disabilities [4]. According to the newspaper El Correo, it indicates that, among the most recurrent needs among blind children and young people, books adapted in braille and computers with the necessary technology are required to allow the student to access the information, being the most required audiobooks.

In [5], using image processing for the translation of the Braille language into the Spanish language, they have a preprocessing stage where they recognize the characters and then separate the information from each cell in order to obtain the 27 letters of the panish alphabet, this research work makes a comparison process with a database where they get the translation of the letters, also saves them in .txt format indicating the saving of the translated data.

*Corresponding Author: Brian Meneses-Claudio, bmeneses@uch.edu.pe

In [6], the research paper presents the design of a prototype of Braille language conversion to English vocabulary and vice versa, being easy to use for the person with visual impairment or blindness, at the moment the device is in the experimental phase, obtaining Good results, but the only drawback is that prior to the use of the electronic device, the person with visual impairment or blindness and the person without it, must be trained to use the equipment.

In [7], the research paper presents an electronic circuit with 6 buttons, this circuit will be able to obtain the Bangla language in text format and then convert it into a Braille code to finally be represented in binary code and thus through the Push buttons provide tactile understanding to people with visual impairment or blindness. In addition to the electronic circuit, it also has a buzzer that will emit sounds for periods of time depending on the radius of the points.

The main objective of the research work is to show the possibility of the implementation of image processing technologies for the translation of the Braille language into the Peruvian alphabet and the possibility of improving the social inclusion of people who have blindness or a disease related to blindness or loss of vision and also the implementation of these systems in educational centers so that these people can adapt to the education system.

Image processing is a set of techniques that are applied to digital images in order to improve or search for relevant information for a particular use. It is currently used for different lines of research such as satellite images in the detection of areas or biomedical for the detection of ills, as well as in this case of the research work in the area of processing and searching for information in an image to be extracted and then interpreted.

The following research work is structured as follows: In section II, the methodology of image processing of the text in Braille will be presented, in addition to presenting the flowcharts of the processes that were carried out to obtain the meaning of each image. In section III, the results of the research work will be shown through a table of accuracy about the tests performed and the percentage of acceptance in each of them. Finally, in section IV, the research work discussions will be presented and the conclusions on the research work will be presented as well as some recommendations for image processing.

2. Methodology

In this part, the steps for the processing of images for the translation of the Braille language into Spanish vocabulary are developed, which consist of the acquisition of the image, then find the centers of each of the characters and then separate them and find the centers and radii of the circles that make up each of the letters in Braille language, these steps are shown in Figure 1 in which at the end you get the letter in Braille format with the identification of the objects that make up the letter in Braille.

2.1. Image Acquisition

For the acquisition of an image, fragments of books were photographed and data on texts in Braille were downloaded from the Internet, then processed. This image must contain the visible points as shown in Figure 2, because it was noted that in many texts only large points are presented and for image processing it is preferable that the format which has all the points.

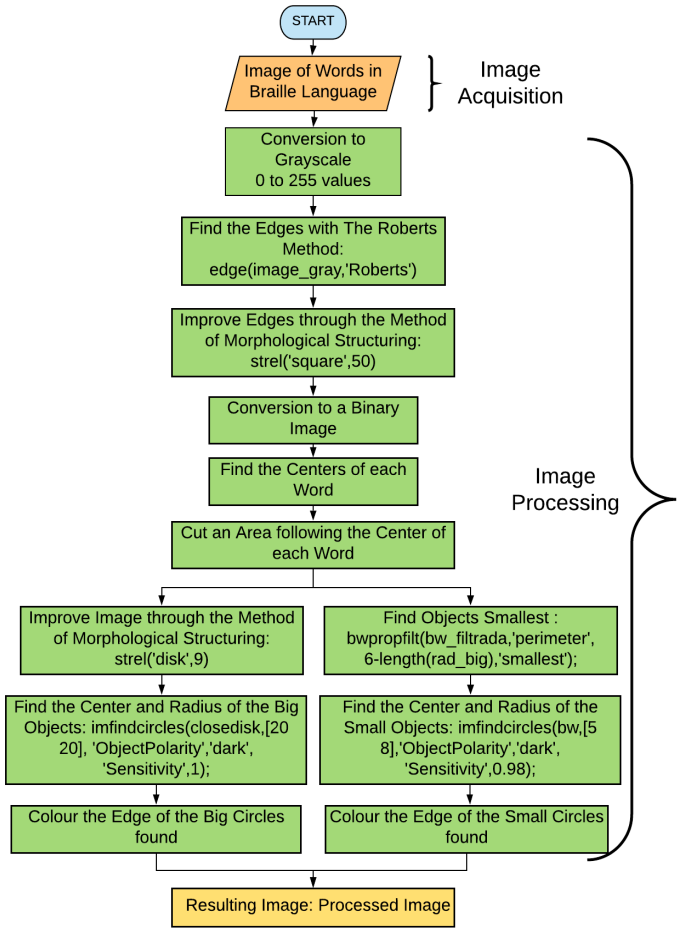


Figure 1: Process Flow Diagram for obtaining the Letter in Braille format and identifying its points.

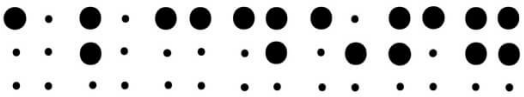


Figure 2: Letters in Braille Format showing large and small points.

In addition, it has to take into account the dimensions of each points because the diameter of the large points must be 1.2mm and the separation of them between each point must be 2.5mm. Each of the letters in Braille must have a separation of approximately 2.3mm.

The scale of the dots in both cases (fragments of books and texts from Internet) are similar because they have the same dimension mention before. Also, these dimensions are international so there is no problem to analyze from different sources.

2.2. Image Processing

At this stage, the image that is entered is in 3 dimensions, so, using the MATLAB software, it will process the image to obtain

the Braille letter with the identification of its points, in addition to the centers and radius of each of them.

For the grayscale conversion, a weighted sum was made taking the values of each pixel that have the image of each dimension and then multiplied with a parameterized value as shown below [8]:

$$0.2989 * R + 0.5870 * G + 0.114 * B \quad (1)$$

After having the grayscale image being this of 2 dimensions as shown in Figure 3 (a), it proceeded to find the edges of the points that have each letter in Braille format, the Roberts Method was used, because it's very easily identifies the diagonal and linear edges, also has the ability to fill spaces [9]. Roberts's method uses a filter that focuses on each pixel following the formula below:

$$\frac{df}{dx} = f(x+1, y) - f(x, y) \quad (2)$$

$$\frac{df}{dy} = f(x, y+1) - f(x, y)$$

Where to locate the pixel (x, y) that is in the ranges from 0 to 255, if the areas have a constant intensity, it will convert them to 0, thus giving the edges of the points.

The next step is to improve the shape of the edges that is why the Morphological Structure Method is used in the form of a square because an area that contains all the points of a letter in Braille language is required, the formula below was followed:

$$\delta_B(X) = X \oplus B = \{x | X \cap B_x \neq \emptyset\} \quad (3)$$

Where, it is indicated that X will go through the whole image, when it passes through B, it will provide the information of the data of the neighbors of that pixel converting to the maximum value of the environment of that neighborhood defined by the element of the structure resulting in Figure 3 (b), where it is verified that within the white objects are the points [10].

The central points are then identified as shown in Figure 3 (c), where the following code was used to identify the "centroids" of the objects:

regionprops(object_image,'centroid')

The main function of it is to verify the sizes of the X and Y axes, and then divide them independently to obtain the midpoint of the object.

The following is to cut each object for a personalized analysis of each letter in Braille format, for this process the centers previously found were used and then estimate the dimensions that each letter can have in Braille format, for this process the following code was used:

imcrop(image_real,'Position',[centroids(1,1)-60 centroids(1,2)-85 120 170]);

This cutting process is done to the image in grayscale because it requires the visualization of the large and small points that have the word in Braille format, as shown in Figure 3 (d) where there are two separate letters and cropped.

The next thing is to find the big points of the cropped image, for that the reader had to binarize the image to eliminate the small points and get to measure and identify the diameter of the large points and colour them red, as shown in the Figure 3 (e), for this process the following code was used:

[center_large,rad_large]=imfindcircles(image_binary,[20 20], 'ObjectPolarity','dark', 'Sensitivity',1);

large_points=viscircles(center_large,rad_large,'Color','r');

In where, it is the center of each circle as well as the radii of each of them.

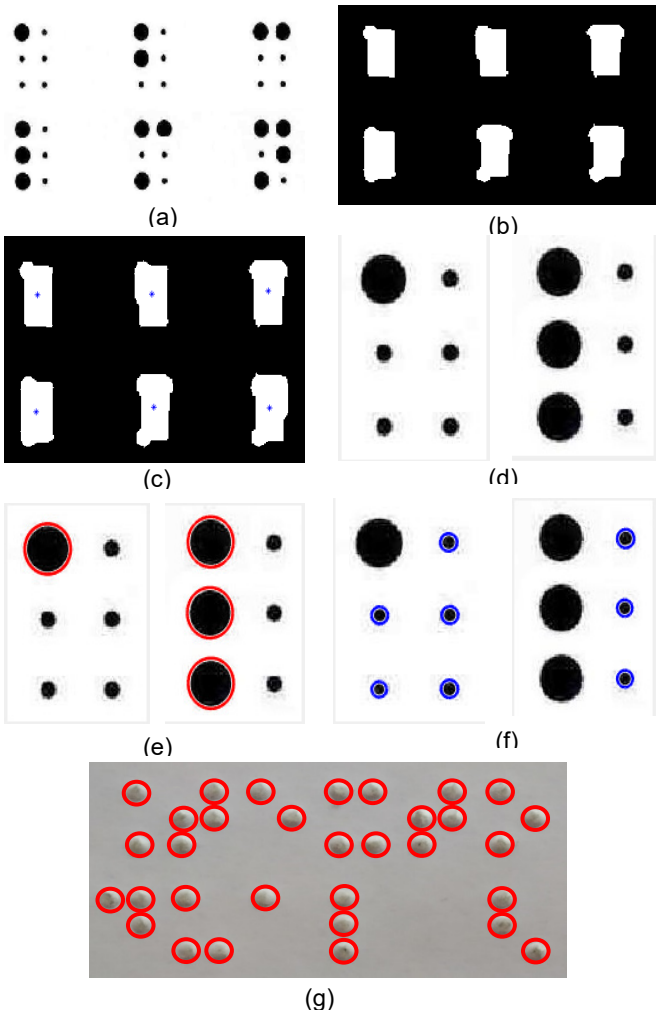


Figure 3: (a) Grayscale Image. (b) Edge detection with Roberts's method and square-shaped enhancement filter. (c) Identification of the Centers of each Object found. (d) Cut out the letters in Braille format. (e) Identification of the Largest Points. (f) Identification of the Smallest Points. (g) Fragment from a Braille Book

- Binarize the image, it means getting the image at 1 and 0.

- Then, use a filter that makes sectored and not eliminate the smallest objects in the image for the following process the following code was used:

bwpropfilt(bw_filter,'perimeter',6-length(rad_large), 'smallest');

The number of large points obtained previously is used and then it subtracts it with 6 which are the maximum points that represent a letter in Braille format, and it indicates that it only require the smallest objects so the software could give the small points. The following code was used to identify the points and their radii:

[center_small,rad_small]=imfindcircles(bw,[5 8],'ObjectPolarity','dark', 'Sensitivity',0.98);

Small_points=viscircles(center_small,rad_small,'Color','b');

The final result is Figure 3 (e), so in this way, the identification of the large and small points in the representations of the words in Braille format is obtained.

This process of identifying the large and small points will serve as how to convert it to a binary matrix and thus differentiate each letter in Braille format by its own identification number.

Table 1: Accuracy of Translation

Class	Predicted class	
	Positive Predictive Value	False Discovery Rate
A	100.0%	0%
B	100.0%	0%
C	100.0%	0%
D	100.0%	0%
E	100.0%	0%
F	100.0%	0%
G	100.0%	0%
H	100.0%	0%
I	100.0%	0%
J	100.0%	0%
K	100.0%	0%
L	100.0%	0%
M	100.0%	0%
N	100.0%	0%
Ñ	100.0%	0%
O	100.0%	0%
P	100.0%	0%
Q	100.0%	0%
R	100.0%	0%
S	100.0%	0%
T	100.0%	0%
U	100.0%	0%
V	100.0%	0%
W	100.0%	0%
X	100.0%	0%
Y	100.0%	0%
Z	100.0%	0%

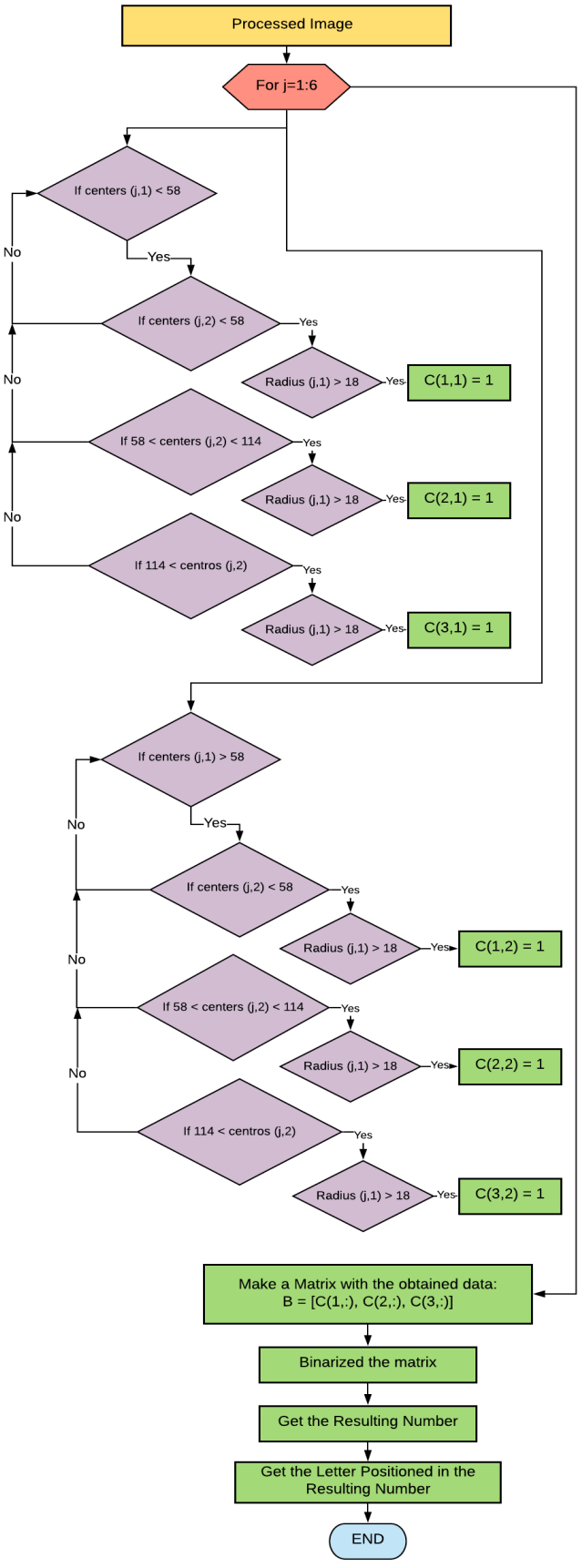


Figure 4: Flowchart of the Conversion of the Image to Matrix and Obtaining Results.

The resulting image will be processed in a loop as shown in Figure 4, which will cover the 6 points that have a letter in Braille, when it detects that there is a large point, it will convert it to a logical 1 and when it detects small points, it will convert to a logical 0 and get a 3x2 matrix. Finishing, this matrix will become binary and will be added, to obtain a number that identifies each letter, and so when it enters images, the algorithm will be able to identify the letter in the Peruvian vocabulary.

For example, when it processes the letter "L" in Braille format, it obtains a matrix, then binarize it in base 2, make an addition with the numbers and finally it obtains the identifying number for each letter:

$$\begin{matrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{matrix} \quad \begin{matrix} 2^0 & 2^1 & 2^2 & 2^3 & 2^5 & 2^6 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{matrix} = 1+0+4+0+16+0 = 21 = L$$

3. Results

Tests were made with 4 images of texts found on the Internet and 2 text from a book in Braille as can see in Figure 3 (g). In each image, each of the letters of the Spanish alphabet was identified, resulting in a 100% percent of translation, as shown in Table I.

These results indicate that the program is capable of translating the Braille language into Spanish vocabulary just by finding the large and small points that have a letter. In addition, it was indicated that the system was made with 6 images with the same resolution, in addition to applying several filters to improve the image because they are very small.

As mentioned above, these results are testing with lowercase Braille letters because a larger matrix format is required for uppercase, which is a matter of adding parameters to the program so the reader can translate uppercase letters in Braille format.

This paper confirms the use of image processing for translate Braille text to Spanish text, also it could record as an audio, and this will be a future work. Also, the process could be inverse with the use of image processing. The main purpose is the translation of Braille letters in Spanish to Peruvian Letters in Spanish.

4. Discussion and Conclusions

The research work confirms the use of image processing for the translation of the letters in Braille format to Spanish letters, in addition, that at the moment tests were made with lowercase letters having their matrix of 3x2.

This type of technology can be applied in educational centers to help the teaching of people with visual disabilities and also train teachers so that they can communicate with students with these deficiencies, therefore, improve social inclusion and study completion statistics by these people.

Braille translation systems is the vast majority tend to be expensive and very difficult access to schools or people who require the system, because of this, the research work is presented to provide open source how to make your Braille language translators.

It is concluded that the binary mode and morphological improvement filters help the image processing because the letters in Braille format are very small and when they are processed, they tend to form distorted pixels giving errors.

It is concluded that to find the large and small points can be done without the need to use an edge analyzer, as was shown in this research work, the reader can use the discard method and indicate that the software searches for small objects and will find them.

It is concluded that each process is independent, so it is necessary to use and assign different variables so that the image processing has an adequate order.

As a work in the future, we want to test the algorithm with books and a webcam to obtain the images in real time, so to know if under movement or image resolution tests the translation changes or errors are generated. Also, attach it to an electronic circuit to make the system portable.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] J. A. A. Pauta, E. P. Vélez and L. Serpa-Andrade, "Braille teaching electronic prototype," 2016 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, 2016, pp. 1-7. https://doi.org/10.1109/ROPEC.2016.7830573

[2] El correo, "la desercion en los estudiantes con discapacidad visual", 2018 [Online]. Available https://diariocorreo.pe/peru/la-desercion-escolar-en-los-estudiantes-con-discapacidad-visual-807248/.

[3] DISCAPNET, "Lenguaje Braille", 2017 [Online]. Available https://www.discapnet.es/areas-tematicas/diseno-para-todos/accesibilidad-de-comunicacion/lenguaje-braille

[4] P. Petrantonakis, V. Kosmidou, M. Nikolaraizi, S. Koutsogiorgou and L. J. Hadjileontiadis, ""SEE and SEE": An Educational Tool for Kids with Hard of Hearing," 2008 Eighth IEEE International Conference on Advanced Learning Technologies, Santander, Cantabria, 2008, pp. 1032-1033. https://doi.org/10.1109/ICALT.2008.45

[5] S. Molina, B. Pérez and J. Gómez, "Literary Braille language translator to Spanish text," 2016 IEEE International Conference on Automatica (ICA-ACCA), Curico, 2016, pp. 1-5. https://doi.org/10.1109/ICA-ACCA.2016.7778514

[6] A. Dasgupta, D. Seth, A. Ghosh and A. Nath, "Real time sign language to Braille interfacing system," 2017 7th International Conference on Communication Systems and Network Technologies (CSNT), Nagpur, 2017, pp. 371-375. https://doi.org/10.1109/CSNT.2017.8418569

[7] N. Sojib and M. Zafar Iqbal, "Single Cell Bangla Braille Book Reader for Visually Impaired People," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, 2018, pp. 1-4. https://doi.org/10.1109/ICBSLP.2018.8554607

[8] Brian Meneses-Claudio, Witman Alvarado-Díaz, Fiorella Flores-Medina, Natalia I. Vargas-Cuentas and Avid Roman-Gonzalez, "Muscles Heating Analysis in Sportspeople to Prevent Muscle Injuries using Thermal Images" International Journal of Advanced Computer Science and Applications(IJACSA), 10(6), 2019. https://doi.org/10.14569/IJACSA.2019.0100606

[9] B. Meneses-Claudio, W. Alvarado-Díaz, F. Flores-Medina, N. Vargas-Cuentas, and A. Roman-Gonzalez, "Detection of Suspicious of Diabetic Feet using Thermal Image," International Journal of Advanced Computer Science and Applications(IJACSA), vol. 10, no. 6, pp. 379–383, 2019. https://doi.org/10.14569/IJACSA.2019.0100648

[10] S. Gnanapriya, P. Esakkipriya, R. Kavipriya, C. Sangeethakamatchi and M. Sandhya, "Identification of organic fruits using color and size features," 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), Chennai, 2017, pp. 160-163. https://doi.org/10.1109/TIAR.2017.8273707

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 408-417 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Resource Selection Service Based on Neural Network in Fog Environment

Nour Mostafa*

College of Engineering and Technology, American University of the Middle East (AUM), Egaila, Kuwait

ARTICLE INFO

Article history:
Received: 14 January, 2020
Accepted: 18 February, 2020
Online: 25 February, 2020

Keywords:
Cloud
Fog
Fog-to-Cloud
Fog-to-Fog
IoT
e2e delay

ABSTRACT

As an emergent technology in Internet of Things (IoT), the ultimate target of fog computing is to provide a widely distributed computational resources and data repository closer to the network edge providing heterogeneous systems both in terms of software and hardware. The fog system must have the capability to deal with huge number of resources and users at the same time, such increased size sometimes presents the issue of performance degradation. Therefore, the fog should be able to support adaptability, scalability, and extensibility to avoid such degradation by adopting efficiently and effectively an optimal resource selection and allocation model. As many fog users have limited interest in, or knowledge of fog middleware issues, it follows that in a large fog environment the best approach would be to have an automated system for resource selection and allocation. Such an approach eliminates the need for user intervention. This paper proposes a fog resource selection service based on neural network to perform resource selection tasks by coordinating with the metascheduler. Five different selection algorithm were used to evaluate the prediction model for resource selection. In addition to introducing a history update and management algorithm to manage and control the storage of the history log records.

1. Introduction

Fog and IoT systems are handling huge numbers of users, resources and tasks. The fog and IoT environment are widely distributed with great diversity, therefore, it is expected that resource selection will be carried out by run time estimates. With the increase size of the fog and IoT and also with its wider acceptance by the 'ordinary' users, some form of automatic resource selection system is desirable. This resource selection system should have the capability to evaluate available resources on the basis of specific criteria as determined by the user.

Because of its huge volume, the fog requires special services for its efficient utilization. With the expansion in the size of the fog and IoT, users are facing new challenges of resource selection to meet their computational requirements without getting into the fine details of available resources. Although, evolving very rapidly and becoming widely accepted, normal users of the fog and IoT find it very difficult to use middleware technologies. Middleware services provide a means to virtualize and aggregate resources. Fog resources are heterogeneous and very complex and require intelligent solutions for resource discovery and characterization. Therefore, in fog environment, the network and fog designers are facing a serious challenge to grant efficient utilization of such huge and widely distributed resources. Such increased size sometimes presents the issue of performance degradation. Therefore the fog system should be able to support adaptability, scalability, and extensibility to avoid such degradation. In order for the fog system to scale well with the increase size of resources and users, we propose a Fog Resources Selection Service (FResS) for fog systems based on neural network that will meet a user's preferences in an optimal response time. The proposed strategy presents a predictive resources selection model to predict and determine the optimal resource to fulfill the user's task. The work introduced in [1] has been extended in this paper where the IoT framework is integrated within the FResS, in addition to testing and analyzing different selection algorithms. Moreover, a history update and management algorithm has been introduced to manage and control the storage of the history log records. This situation requires that the fog user should be provided with a helper service to facilitate resource selection. Resource selection can be done by taking into consideration either, cost, time, cost and time, performance, or any other constraint. This resource selection service can have its own resource selection algorithms and can also be deployed at varying distribution levels. Currently the majority

*Corresponding Author: Nour Moustafa, nour.moustafa@aum.edu.kw

of metaschedulers are deployed at the cloud level which will be a bottleneck for larger fog and IoT systems. A higher distribution level will alleviate this bottleneck. The future fog and IoT is also intended to be more user focused.

The advanced fog selection for task execution is enabled by the resource selector which uses the FResS to make a rum time prediction to select the best host/resource. Due to the resource-limited fogs, the proposed model avoid incurring overhead when executing the task. In addition to decreasing the overall end-to-end (e2e) latency of the system. The proposed FResS model select resources based on criteria mentioned by the user, which provide QoS levels that meet the users' expectations. Experimental performance evaluation shows an overall improvement in terms of bandwidth consumption and task execution times, in addition to maintaining adequate QoS levels.

The reminder of the paper is organized as follows: Section 2 presents background reading and related work. To provide a deeper understanding of this area, section 3 introduces the fog prediction and selection scheme. Section 4 presents the multi-level load balancing technique. Section 5 presents the evaluation results using workloads and simulations respectively. Finally, the paper concludes with section 6, focusing on the contribution along with future recommendation.

2. Related Work

The emergence of cloud, fog and IoT approaches as a comparatively new approaches to distributed computing, which has become popular in the last few years. These approaches addresses the problem of organizing large scale computer societies, called Virtual Machine, which are able to use and share large sets of distributed resources (e.g. computational resources and data repositories) among them. With the emergence of several cloud and fog service providers, it is expected that system performance and run time estimates will be used for resource selection, which make organization to move from one service provider to another to meet their needs. Predictions are a very important step towards automatic resource management [1][2]. As the capabilities of IoT devices are limited, more attention is required when allocating tasks on these devices to assign the workload optimally on available resources.

In [3] authors aims to enable parallel execution by splitting the allocated services on the available resources. To achieve low delay on service allocation the proposed model distributed/matched the desired services among the available Fog-to-Cloud (F2C) resources. The author in [4] proposed a resource allocation strategy for resource selection by adopting the Price Timed Petri Nets (PTPNs) which allow the user to select the resources autonomously based on the price cost and time cost to complete a task. In addition to sorting the resources into groups based on their credibility, then the users will be assigned to different groups because the credibility that they valuate the resources is different. The authors in [5] proposed an online algorithm for task allocation in mobile IoT networks by considering two scenarios to address the scenario of tasks information is unknown and incomplete local task information, respectively. The task allocation decision is done based on the current network status, energy consumption in order to satisfy the mobile device battery capacity constraint and to minimize the energy consumption of the current location.

In [6], the authors proposed an online task scheduling scheme called FairTS by adopting and combining the deep reinforcement learning (DRL) and dominant resource fairness (DRF) techniques. The proposed scheme views the agent's policy, take the current state using neural network to learn from experience, and outputs the action selection probability vector. The agent observe the resource availability, pick a task, and decide its fog resource allocation based on a resource availability that meet the task completion deadline to achieve resource fairness among tasks.

The author in [7] proposed offloading scheme, where the task data submitted by the user is uploaded to its nearby fog node, then the fog node will take a decision for offloading the task to its neighbor fog node or remote cloud to satisfy the delay deadline received from the end user. In [8] the authors used the machine learning as a prediction model to predict cloud processing resource consumption. The proposed model uses an artificial intelligence algorithm for future request prediction by dividing the incoming request into multiple consumption classes depending on their consumption amount, then the request is distributed among processing unites. For an incoming request, the adopted algorithm is used to predict the resource amount consumed to complete the task, a placement algorithm is used to allocate it to the best processing unit.

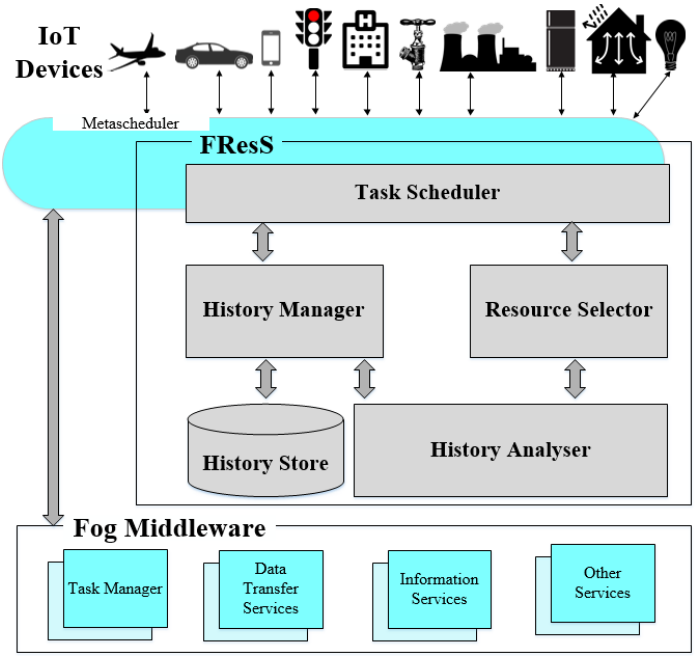
Although there exists research investigating the resources selection and allocation issues in fog computing, however it is still premature. As mentioned earlier, this project resulted in the development of two novel components: a run time prediction model and a distributed architecture of a Fog Resource Selection Service (FResS), which provides a degree of autonomy in determining the optimal resources to execute a particular task.

3. Fog Prediction and Selection Scheme

In our paper [1] a novel approach has been introduced in fog computing, namely, a Fog to Fog (F2F) resource selection algorithm called FResS, that provide automatic resource selection system. This resource selection system have the capability to evaluate available resource on the basis of specific criteria as determined by the user. The idea was to develop a new technique which require no user input. This work resulted in a run time prediction model based on historical execution logs and a fog resource selection service. The main modules of the FResS framework are: Task Scheduler, Resource Selector, and History Analyzer. These modules perform different resource management functions, the most obvious of which is resource prediction and selection. All of these modules are discussed separately in the coming sections.

A detailed architectural diagram of FResS is shown in Figure 1 highlighting the fact that FResS works along with the metascheduler. The FResS architecture was developed by keeping in mind its distribution level, which can be either IoT level, or fog level. The recommended distribution level is the IoT level, if deployed at the IoT level, FResS stores historical data related to the single IoT, it makes the management of execution logs straightforward because the overall volume of data will be very small. It will also provide fast and accurate predictions, again

because of smaller and related data. If deployed at the fog level it can become a single point of failure and a performance bottleneck. Implementation details will be very much similar for the different levels of distribution.



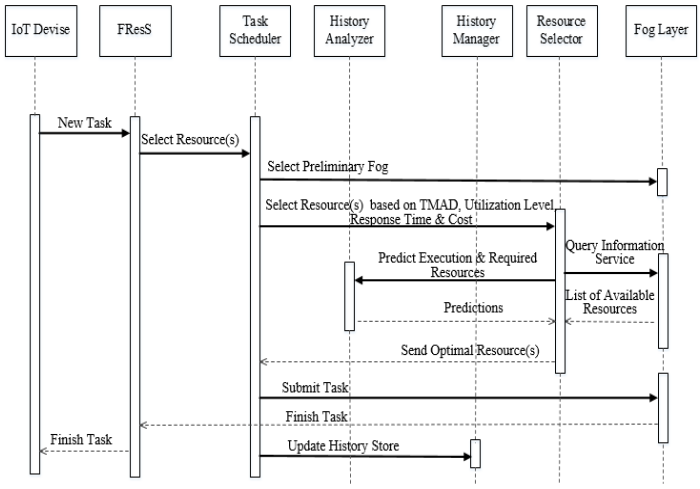
architecture

3.1. Task Scheduler Module

This is the main component interacting with the fog middleware through a metascheduler. Its implementation depends on the underlying fog middleware. Depending on its distribution level, whether it is IoT level or fog level, the Task Scheduler Module can have different implementation details. At a minimum it provides interfaces between the metascheduler and the FResS components, but it can also act as a metascheduler in itself. For evaluation purposes a simplified Task Scheduler Module was implemented by removing all external constraints.

The Task Scheduler sits between an IoT devise/user and the fog middleware. It provides interfaces to the IoT, the fog middleware, and other modules of the metascheduler. The sequence diagram shown in Figure 2 explains the sequence of events during a typical task submission process. The Task Scheduler Module and the metascheduler module coordinate with the lower layers of the fog to execute different tasks. Lower layers of the fog provide basic services i.e. data transfer, task submission, task management, and information services. Once a task is submitted by the user to the Task Scheduler, it will be connected to a preliminary fog based on users' location, and then the Task Scheduler Module forwards it to the Resource Selector Module.

The configuration plays a key role in selecting resources for each task and predicting its execution time. If the task selector returns more than one resource for a particular task, then an optimal resource will be selected by the Task Scheduler to fulfill users' request based on utilization level, which allow the users' preference to be considered while assigning the task to resource(s) which is more suitable for large scale domain such as fog computing. The progress of the tasks and resources will be constantly monitored by the proposed model, which take decisions based on that. The History Manager will receive the execution logs from the Task Scheduler after every execution, and subsequently predictions will be made for future tasks using these logs. The sequence diagram in Figure 2 shows the interaction among the different modules of FResS during a typical task submission process.



FResS model

3.2. Resource Selector Module

The Resource Selector Module analyses the workload of the incoming task to provide automatic resource selection by using the run time predictions. The Resource Selector depending on the configuration, along with their run time predictions, will forward a list of selected resources to the Task Scheduler. The Task Scheduler Module returns the resource list to the metascheduler to complete task submission. The resource selection on the fog is a multi-step process. The first step tests the eligibility of the discovered resources against the essential QoS criteria, e.g. privacy, memory, cost, availability, and delay. The list of the selected resources should be capable of executing the current task if there is no deadline constraint. Step one is performed by the metascheduler on its own.

In step two, the ability of resources to meet deadlines will be evaluated by evaluating run time predictions. For optimum task allocation, these predictions are then utilized by the metascheduler. To make optimum resource selections the metascheduler also considers additional constraints of user preferences, data locality, co-allocation, cost, workflow constraints, and advance reservation. These additional constraints are also evaluated by the metascheduler after receiving a ranked list of resources along with predictions from FResS. Predictions help the metascheduler to make a decision on the required duration of advance reservation. The co-allocation decision is facilitated with this predicted knowledge of expected run times.

The Resource Selector Module receives its input from the Task Scheduler Module along with the currently available resource list. It forwards incoming tasks to the History Analyzer Module to generate run time predictions for the available resources. The History Analyzer Module then uses algorithm 6, discussed in the next section, to generate run time predictions. For the evaluation

of FResS an experimental strategy was devised in which five algorithms 1, 2, 3, 4, 5 were used by the Resource Selector Module for resource selection. These algorithms are named as history, static, random, history plus static, and history plus random.

Data: workload history
Input: available resource list
Input: task submission description file
Result: selected resource
initialization;
getPredictions(new task submission description file, available resources);
selected resource is one with the minimum run time prediction;
return selected resource;

Algorithm 1: Selecting resource using history

Data: workload history
Input: available resource list
Input: task submission description file
Result: selected resource
initialization;
getPredictions(new task submission description file, available resources);
selected resource is one with the highest mips rating;
return selected resource;

Algorithm 2: Selecting resource using static performance capability

Data: workload history
Input: available resource list
Input: task submission description file
Result: selected resource
initialization;
getPredictions(new task submission description file, available resources);
select a resource randomly from the available resource list;
return selected resource;

Algorithm 3: Random resource selection

Data: workload history
Input: available resource list
Input: task submission description file
Result: selected resource
initialization;
getPredictions(new task submission description file, available resources);
select 10 resources with the least run time predictions;
from these resources pick resource with highest mips rating;
return selected resource;

Algorithm 4: Selecting resource using history and mips rating

Data: workload history
Input: available resource list
Input: task submission description file
Result: selected resource
initialization;
getPredictions(new task submission description file, available resources);
select 10 resources with the least run time predictions;
from these resources pick any resource randomly;
return selected resource;

Algorithm 5: Selecting resource using history and random selection

Selecting the fastest resource depending on its static number crunching capacity is a straightforward solution to the resource selection problem. It is approximately comparable to the Condor matchmaking technique [9]. It is mentioned above as static resource selection, which suffers from the problem that it does not have a feedback mechanism to inform the system about any anomaly. The history based resource selection on the other hand has a feedback mechanism to rectify any anomaly resulting from changed resource capability.

The random resource selection results in an even distribution of tasks among candidate resources without considering the difference in their capabilities. On heterogeneous fog system, it is not a desired outcome where high power resources are required to get more tasks. The random resource selection also does not provide any estimates of expected execution time, leaving the user in a 'wait and see' situation. The historical resource selection is combined with last two algorithms, 4 and 5 to overcome the drawback of static and random resource selection. Another shortcoming of the historical resource selection is the uneven load distribution similar to the static resource selection. However, the presence of a feedback loop shown in Figure 4 reduces the severity of this problem. The performance of all these algorithms will be evaluated in detail in the simulation result section. Once similar tasks are found, run time predictions are generated for all resources in the list. The finished list is returned to the Resource Selector Module.

3.3. History Analyzer Module

The incoming task along with the list of available resources is forwarded to the History Analyzer Module to generate run time predictions by examining through the history execution logs to find similar tasks using algorithm 6, this process is conducted using Artificial Neural Networks (ANNs) [10]. Once similar tasks are found, run time predictions are generated for all resources in the list. The complete list is then returned to the Resource Selection Module. ANNs are suitable for training over hundreds or even thousands of passes of data sets[11]. The increased size of the fog and IoT systems brings new challenges of scalability and extensibility and the need for automated processing becomes clear therefore, accurate results can be achieved. The design of ANNs generally comprises three layers, input, middle (or "hidden"), and output layers [10]. The input layer is the starting point where the data enters the system. The input data are passed to the hidden layer for processing which is the intermediate processing unit, and then passes the new signal onto the output layer. The learning phase plays a key role in neural network for processing information

using the interconnection weights. The strength of the input data is measured by weights, therefore, the process of finding the best set of weights for the neural network is referred to as training or learning. In order to get accurate results, the weights are modified at the input layer by passing input values, and then the network's predictions will be measured to check how the predictions' values are close to the training sets. In addition, the accuracy of the prediction results is compared to the actual system results as shown in the simulation result section 5, Figure 3 shows a single unit network.

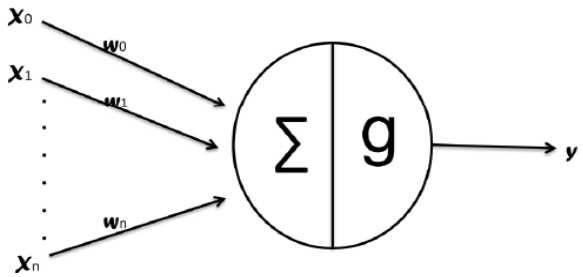


Figure 3: An Artificial Neuron Unit [12]

An amount proportional will be changed to the difference between the desired output and the actual output to achieve weight adjustment [13]. The general mathematic definition [14] is as shown in (1) describes the relationship between the input and output mapping:

$$y(x) = g(\sum_{i=0}^{n}(w_i x_i)) \quad (1)$$

the input of each neuron is represented by x with (n+1) input dendrites (x_0….. x_n) and the axon $y(x)$ computes the output. In order to supply a node activation the weighted signals are initialized with (w_0….. w_n). The activation function is the identity function and is represented by g which computes the weighted sum of the inputs from other neurons and outputs it. The aim is to be able to produce the output within desired accuracy matching the input pattern [15].

After the historical execution logs are examined by the History Analyzer to generate estimates for the required resources by finding similar task with sufficient accuracy, once predictions are generated, the complete list will be forwarded to the Resource Selector to determine the optimal resources to execute the task. Algorithm 6 forms the core of FResS, which is used to generate rum time prediction for new tasks on the available resources.

The process of identifying the resource id dependent on five parameters: delay, price, previous execution time, memory requested, and submission time. Workloads can be modelled at any level but it is found that modelling at a user level is more realistic, this is because users of fog networks tend to repeat same tasks using the same data [12] [16]. Therefore, the results are interesting since prediction of required resources of the new task will be made using past execution parameters.

The proposed algorithm combines the static and random algorithms with the historical resource selection to overcome their disadvantages, e.g. in certain situations there is a possibility of selecting the same resource repeatedly by the Task Scheduler, creating an uneven load distribution. Consequently, the presence of the feedback loop as shown in Figure 4 overcome the severity of this problem.

Data: user preferences and task log history
Input: available resource list
Input: new task submission description file
Result: run time prediction for the resource list
Start
While Taskset <> empty
do
  process request for Task(s) k_i;
  if (Ki submitted by new user/device) then;
    Resource Selector send available resources to Task Scheduler;
    Task Scheduler send user preferences to Task Manager;
    if (Task Manager found optimal resource(s));
      send back optimal resource(s) to Task Scheduler;
      execute k_i on the optimal resource;
  else if (Ki submitted by existing user/devise) then;
    NN predict required resources for k_i;
    Send back predicted resource(s) to execute k_i;
    execute k_i on the predicted resource;
  else
    execute k_i on the preliminary fog;
  update Task Manager log history;
  select next task;
  end
end

Algorithm 6: Making predictions for run time

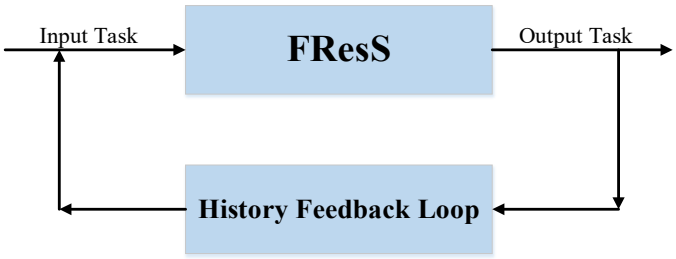


Figure 4: History as a feedback loop in the scheduling process

On the other hand the history management algorithm (algorithm 7) works separately and is responsible for keeping the history up to date. It means that it stores new history records after successful executions and removes older history records which is updated continuously after every execution and subsequently used by ANN to make predictions for future tasks. Subsequently, the increasing number of the historical executions records will increase the accuracy of predictions [17][18].

3.4. History Manager Module

The History Manager is interfaced with the Task Scheduler and history database, it is responsible for keeping the history data up to date. History updates are performed by using algorithm 7. The History Manager Module controls the storage and removal of

execution records as required. Algorithm 7 performs history management by storing completed tasks data in the database as soon as they are received. After this step, a query is sent to the database to retrieve the size of the node related to the new task, just stored. If the size of this node is less than, or equal to, the maximum history limit then no action is taken. Otherwise, if the size of this node is larger than the maximum, then the oldest member of this node is deleted. This module gathers different QoS data e.g. delay, cost, previous execution time and memory requested etc.

Data: workload history
Input: execution log of completed job
Input: task submission description file
Result: history updated
initialization;
store record to the history database;
if size of this cluster is more then the size limit then
  delete the oldest record;
end

Algorithm 7: Update history

3.5. Complexity of Algorithms

Algorithms 6 and 7, presented above, form the core of FResS. The predictions algorithm (algorithm 6) is used to generate run time predictions for new tasks on the available resources. On the other hand the history management algorithm (algorithm 7) works separately and is responsible for keeping the history up to date. It means that it stores new history record after successful executions and removes older history records.

The complexity of predictions for algorithm 6 is of $O(n)$ where n is the size of available resources. On the other hand, complexity for history management algorithm 7 is of the $O(1)$. These algorithms will become part of the task submission service. The history management algorithm 7 can be executed offline, after the task is finished, so it's speed is not very critical to the overall execution time because it is not contributing to the overall delay. Algorithm 6, on the other hand, will contribute to the total delay in the execution time.

3.6. Fog Workflows and FResS

Predictions are utilized by the task management system to select resources to execute incoming workflows consisting of many tasks. These workflows can be very simple, consisting of only one task, as shown in Figure 5, or can be very complex, consisting of many interdependent tasks, as shown in Figure 6. These tasks can have dependencies which restrict them to execute in a certain sequential order or execute concurrently. Predictions for each task within a workflow are generated separately.

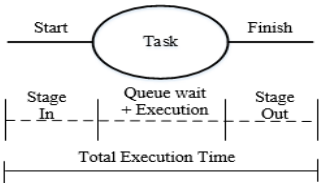
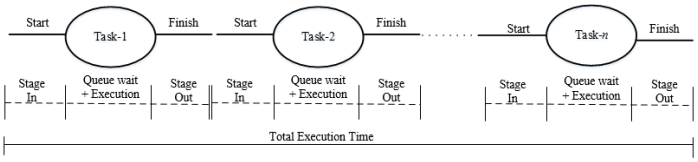


Figure 5: Workflow consisting of only single task



Quality of Service (QoS) refers to the level of performance and service that a given user will experience at a certain point in time, when starting a certain operation on a certain instance of a service. QoS support refers to the possibility of a certain level of required performance being available from a certain resource or not. QoS is very important to fog applications which are run collaboratively in real time. QoS requirements can be mentioned in terms of cost, delay, scalability, fault tolerance, etc.

Normally QoS requirements for these workflows are applied to the entire workflow. To achieve the best timing performance it is recommended that the workflow is partitioned in such a way that parallel activities finish at the same time. Because of the global nature of the fog, locality will also play its role to achieve best QoS results. Sometimes if the user wants to see intermediate results, then QoS can also be mentioned at the task level. Achieving QoS either at task level or workflow level boils down to the task or activity level. To achieve QoS at the workflow level requires that tasks are managed properly. Incorrect management of these tasks can damage QoS at a very high level.

The proposed FResS model is responsible for the overall execution of workflows. These executions, in certain situations, may also require advance reservation and co-allocation. The FResS performs its task comparatively easily when accurate run time predictions are available for tasks within a certain workload. Advance reservation is a technique in which resources are reserved for a certain task to start at a certain time and execute for certain duration. On the fog, advance reservation will be expensive so its accuracy is of prime importance. With the predictions already available, the FResS can reserve resources for accurate time duration, resulting in an efficient usage of resources.

4. A Multi-Level Load Balancing technique in Fog Computing

Internet traffic growing exponentially and congested with thousands and millions of devices [20], quality services should be provided with higher availability by the service provider. With the increase number of services in fog and IoT computing the network load is expected to increase which will create a new challenge to the researchers and network designer. The aim is to achieve better performance in terms of the total time to execute the tasks load balancing to accomplish better QoS, in addition to increasing the number of task acceptance. The fog computing attempts to integrate multiple distributed and heterogeneous resources, which are normally under separate administrative domains resulting in under-utilization or over-utilization. A multi-level load balancing approach has been proposed to provide better load balance. As mentioned above in algorithm 6 user preferences will be considered for the incoming task. Therefore, satisfying the below metrics and parameters optimally will led to achieve better performance of the fog system.

4.1. Metrics

- Cost: IoT and fog users are interested in the best performance at the lowest cost, but on the other hand the resource owners are interested in the overall system throughput. Therefore, the efficient algorithm should consider the cost parameters by scheduling the incoming task to a resource based on the cost.
- Response time: is the total execution time taken by the task to be completed, it starts from submitting the task by the user or device and end with receiving the completed task from the service provider. A smaller response time is always desirable.
- Delay: Researchers and service providers stress on the importance of guaranteeing acceptable delay, since this is an important metric in terms of QoS, the proposed model adopted the tenant maximum acceptable delay (TMAD) [21]. TMAD sorts the incoming tasks to a higher priority task or lower priority by considering different prioritization of users based on their important for the system according to the subscription plan each user pay for.
- Scalability: bottlenecks in fog computing system arises from the huge number of users, resources, and services, therefore to be able to support scalability, and extensibility, this is an important metric which determine if the system is able to achieve better load balancing with a restricted number of resources.
- Fault Tolerant: the larger the system, the more frequent failures result, IoT and fog systems should provide the ability to perform correctly in case of failure by having a backup

5. Simulation Results

The performance of the proposed solution is evaluated in this section by carrying out some experiments using CloudSim [21], selected for its flexibility and wider acceptance by the cloud and fog community. The proposed model performance was evaluated by comparing it with existing scheme and highlighting the differences and advancement. IoT devices were simulated as geographically distributed nodes. The fog network represented as a graph and storage sites were modeled as a set of nodes. A computing power, storage capacity, narrow and broad bandwidth, and memory have been configured to all nodes to be close to reality. The number of tasks, and capacity of storage nodes were varied to show the complexity of the proposed model and simulate different scenarios. The simulator defined 2 cloud sites, 5 fog sites, and 150 IoT devices, 5000 task requests, and 500 to 2000 mb/sec the connectivity of the bandwidth.

Scheduling with predictions helps the scheduler to select the best resource(s) for a certain task and it can be compared with two other techniques i.e. random and static. In a random selection the scheduler selects any resource from the qualifying resources. Qualifying resources are those which meet the minimum requirement criteria for a certain task. On the other hand static selection is one in which the resource with the highest number crunching speed is picked. In a real world scenario, number crunching speed can be based on any popular benchmark. Experimentation was conducted with CloudSim, which defines resource capability in term of Million Instructions Per Second (MIPS), hence in the static resource selection, resource with the highest MIPS rating will be selected from the qualifying resources.

In a deadline based scheduling scheme the function of the scheduler is to meet a deadline specified by the user. It is quite possible that more than one resource can meet this deadline, established by comparing the deadline with the predicted time. In such cases the scheduler can select the fastest resource, which is referred to as history based selection. It is expected that history based resource selection will create an uneven load distribution on the fog system. There are two other options: (1) select the fastest resource from the selected fastest resources, (2) select any resource randomly from the qualifying list. These options were given the name of history plus static and history plus random.

As mentioned in section 3.2, it can be seen that in all, there are five scheduling algorithms, i.e. History, Random, Static, History plus Random, and History plus Static. These algorithms were evaluated using simulation. There are two aspects of the performance of the scheduling algorithms: (1) the execution of tasks within a minimum possible time, (2) the distribution of loads evenly among the available resources. The first experiment was designed to compare the performance of these algorithms in terms of the total execution time to complete the same tasks. In this case the same tasks were submitted to the fog by using each of the five algorithms, one by one, and the total time to complete these tasks was recorded. This experiment was repeated for the tasks created by all workloads, one by one. Results were plotted in the form of bar graphs for each workload based tasks, separately, as shown in Figure 7.

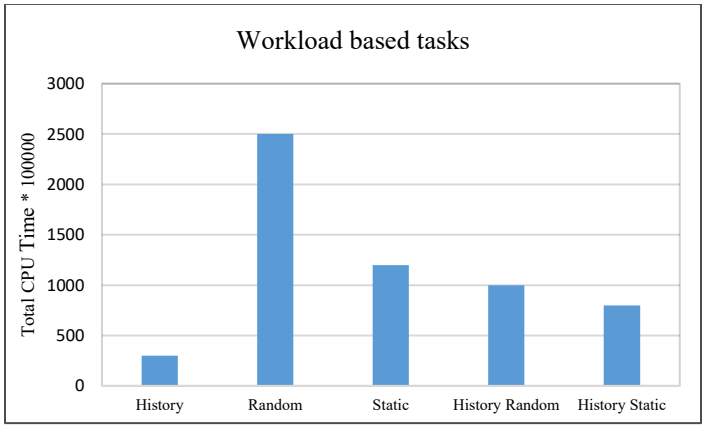


Figure.7 Comparison of total CPU time to execute 5000 tasks for different resource selection algorithms

Different scheduling algorithms were used to evaluate the performance of the predictor i.e. history, static, random, history plus static and history plus random. It was seen that overall execution time was minimum for the static and history based scheduler but it also created issues of load imbalance. History plus random and history plus static provided better load balance, but took larger overall time to finish the same tasks. It was concluded that selective task scheduling will create a load imbalance but a history based solution will operate to fix this imbalance because of the presence of a feedback loop. Random and static schedulers are missing this feedback loop, and hence lack the self-healing property of history based algorithms. Overall performance of FResS was found to be better when using the historical technique, or any hybrid technique that incorporated the historical technique

The accuracy of predictions was also compared with user provided run time estimates. Users can provide run time estimates while submitting tasks to the fog or IoT device. User run time estimates are recorded by all workloads. Given that predictions are

made, the next natural step is to determine the accuracy of these predictions. The accuracy of predictions is calculated by using (3) by the Performance Evaluator.

$$accuracy = \begin{cases} 1 & \text{if} \quad P_{RT} = R_T \\ P_{RT}/R_T & \text{if} \quad P_{RT} < R_T \\ R_T/P_{RT} & \text{if} \quad P_{RT} > R_T \end{cases} \tag{3}$$

Where, P_{RT} is predicted run time and R_T is actual run time.

The accuracy of predictions achieved by both the user and the system is presented in Table 1.

Table 1: Comparison of user's prediction accuracy and system's prediction accuracy

# No. of tasks	Average User % Accuracy	Average System % Accuracy	System 90% Confidence Interval	System 95% Confidence Interval	User 90% Confidence Interval	User 95% Confidence Interval
1000	5.9	73.0	81.4	83.6	2.0	2.0
2000	9.8	79.8	84.7	87.9	8	10
3000	15.3	86.9	88.8	90.7	14.7	16.5
4000	20.0	90.2	91.9	93.3	20.8	23.4
5000	27.1	95.0	96.2	96.9	35.0	39.0

From the results presented in Table 1, it can be seen that the smallest accuracy of 73% was achieved for the workloads of 1000 tasks. The largest accuracy of 95% was achieved for the workloads of 5000 tasks. The remaining workloads have also shown very high accuracy. The accuracy of predictions was also compared with user provided run time estimates while submitting tasks to the fog. User run time estimates are recorded, and as shown in the results in Table 1, the accuracy of user run time estimates lies between 5.9% and 27.1%. It can be concluded therefore that user accuracy is very poor when compared with the achieved accuracy from the predictor.

Another metric of interest, in the fog context, is the number of tasks falling within a certain confidence interval. This interval is defined by the Chebyshevs inequality theorem [22] and can be calculated when a sample's mean is being used as a predictor. On the fog where deadline based scheduling is expected, this confidence interval gives a better idea to the user about the expected task completion time. Chebyshevs' theorem states that the portion of data that lies within k standard deviations to the either side of the mean is at least $1 - \frac{1}{k^2}$ of any data set [22] [23] and it can be expressed as (3):

$$c = \left(1 - \frac{1}{k^2}\right) * 100 \tag{4}$$

Where c is referred to as the confidence interval and k is the number of standard deviations. The Performance Evaluator calculates the predicted tasks falling within the 90% confidence interval and 95% confidence interval. The results of both of these calculations are shown in Table 1. The predictor performed better again, since 81.4% to 96.2% tasks were completed within a 90% confidence interval and 83.6% to 96.9% tasks were completed within a 95% confidence interval. On the other hand, for user run time estimates, only 2.0% to 35.0% tasks were completed within a 90% confidence interval and 2.0% to 39.0% tasks were completed within a 95% confidence interval.

Large number of tasks were executed, first using the existing scheme then the prediction FResS model. The prediction model were generated using queries logs of large number of tasks from CloudSim to evaluate the performance of the proposed model, resource predictions from ANN were generated and used to execute the incoming tasks. The time taken by FResS and the current scheme to execute tasks was measured.

The stages of executing tasks on the fog are explained in the below section, which start with submitting the task and end with the retrieving results after completion. These stages comprise four major components and are given below:

1. **Stage in**: transferring the task(s) to executable resource.
2. **Waiting time**: waiting in the queue for its execution turn.
3. **Run time**: the task is assigned for execution, which marks the start of the run time.
4. **Stage out**: transferring the completed task(s) to the originating node.

Stage in also called stage one, starts with transferring the task from its originating node to the executing node. Once the destination node receives all the required data, then the task is forwarded to the local task manager, which in turns puts it in the waiting queue. The queue wait time represent the time spent in the queue for its execution turn, it depend on the load of the executing node. The run time in stage three marks the start of task execution. In stage four after execution is completed, the task is returned to the originating node which is called the stage out phase. The total turnaround time is given in (2).

$$T_{TRT} = T_{SIT} + T_{QWT} + T_{RT} + T_{SOT} \tag{2}$$

where:

T_{TRT} = *Total Run Time*

T_{SIT} = *Stage In Time*

T_{QWT} = *Queue Wait Time*

T_{RT} = *Run Time*

T_{SOT} = *Stage Out Time*

To evaluate FResS a number of experiments we designed. The purpose of these experiments was to evaluate overall time performance by varying the number of tasks. The tasks response time outlined in Figure 8 show that the FResS model outperform the existing scheme, and the results were promising where the tasks response time was decreased by 33%, as shown in Table 2 showing significant time savings.

Since the fog users have a limited network bandwidth to communicate among the system, such a high communication cost

is not acceptable. In such environment, nodes within a region is provided with broader bandwidth whereas nodes across region is provided with narrow bandwidth. Since many nodes from different regions try to connect to each other through a narrow bandwidth, causing congested network traffic which in turn will affect the task execution time which is an important factor, therefore, carrying out some experiments by varying and expanding the network bandwidths was important in order to simulate the real world environment, the bandwidth between sites was varied in a set of experiments to evaluate the performance of both models.

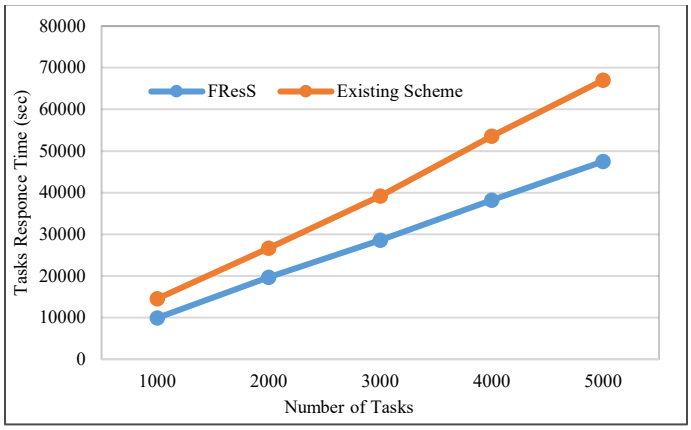


Figure 8: Comparative analysis of Response Time with no of tasks

Table 2. Simulation results: Task Turnaround Time using FResS and Existing scheme

# No. of tasks	Task Turnaround Time using existing scheme (Sec)	Task Turnaround Time using FResS (Sec)	Difference (Sec)
1000	14540	9920	4620
2000	26703	19710	6993
3000	39243	28608	10635
4000	53600	38214	15386
5000	67020	47512	19508
Total	201106	143964	57142
Average	28792	40221	33%

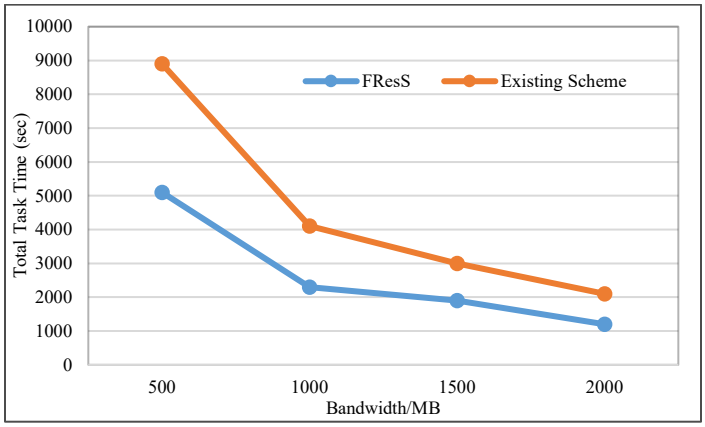


Figure 9: Total task time with varying bandwidth

Figure 9 shows the result of sets of experiments that were run on both narrow and broad bandwidth, as shown in the results, the FResS outperforms the existing scheme, even though setting broader bandwidth decreases the differences of task execution time, however, the difference is still significant in all scenarios. Consequently, it can be concluded that the cost and resource utilization are effectively considered in the proposed FResS model and outperform the existing scheme.

6. Conclusion and Future Work

to overcome the drawback of the cloud computing, the fog computing has emerged with the concept of sharing computational resources and information services by offloading the resources to the edge network to be closer to the devices that originate the requests rather than transporting the tasks to the distant cloud, thus reducing communication overhead, bandwidth consumption and latency. However, the complexity and dynamic nature of the fog systems demand a more coordination platform to handle such a large-scale numbers of resources, users and tasks requests. Therefore, the existing approaches, however, often exhibit a high cost in terms of response time and bandwidth consumption. These shortcomings were overcome by presenting a fully distributed FResS, which stores historical data related to a single user, which simplifies tasks management.

This paper proposed an extension to the resource selection service FResS based on neural network. The proposed prediction model minimize the total overhead of the task turnaround time of the incoming task by predicting and informing the system the required resources, and the database are updated constantly by storing the result of NN tool. The proposed prediction model is simple and shows comparatively lower overheads and provide high possible accuracy.

An evaluation strategy was devised to test five different algorithm used by the Resource Selector Module to provide a detailed performance analysis for each one. A History update algorithm was introduced to manage the history execution logs, as the size in IoT and fog environments are very limited. In addition, the experiments showed that the proposed prediction model distribute the load based on the user preferences. Another major strength of the proposed model is its effective resources utilization achieved by decreasing overall cost, response time, and bandwidth usage. In the future work, more metrics will be added to improve the load balancing, in addition to distributing the proposed model on the cloud layer. Moreover, more experiments will be carried out to define the size of the history records.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The writer would like to thank the American University of the Middle East for its support.

References

[1] N. Mostafa, "Cooperative Fog Communications using A Multi-Level Load Balancing," 2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC), Rome, Italy, 2019, pp. 45-51.doi: 10.1109/FMEC.2019.8795325.

[2] S. Zahara, I. Pratomo and D. S. Rahardjo, "Application and data level interoperability on virtual machine in cloud computing environment," 2015

1st International Conference on Wireless and Telematics (ICWT), Manado, 2015, pp.1-5. doi: 10.1109/ICWT.2015.7449238.

[3] V. B. Souza, X. Masip-Bruin, E. Marin-Tordera, W. Ramirez and S. Sanchez, "Towards Distributed Service Allocation in Fog-to-Cloud (F2C) Scenarios," 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, 2016, pp. 1-6. doi: 10.1109/GLOCOM.2016.7842341.

[4] L. Ni, J. Zhang, C. Jiang, C. Yan and K. Yu, "Resource Allocation Strategy in Fog Computing Based on Priced Timed Petri Nets," in IEEE Internet of Things Journal, vol. 4, no. 5, pp. 1216-1228, Oct. 2017. doi: 10.1109/JIOT.2017.2709814.

[5] J. Yao and N. Ansari, "Task Allocation in Fog-Aided Mobile IoT by Lyapunov Online Reinforcement Learning," in IEEE Transactions on Green Communications and Networking. doi: 10.1109/TGCN.2019.2956626.

[6] S. Bian, X. Huang and Z. Shao, "Online Task Scheduling for Fog Computing with Multi-Resource Fairness," 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall), Honolulu, HI, USA, 2019, pp. 1-5. doi: 10.1109/VTCFall.2019.8891573.

[7] M. Mukherjee et al., "Task Data Offloading and Resource Allocation in Fog Computing With Multi-Task Delay Guarantee," in IEEE Access, vol. 7, pp. 152911-152918, 2019. doi: 10.1109/ACCESS.2019.2941741.

[8] F. Derakhshan, H. Roessler, P. Schefczik and S. Randriamasy, "On prediction of resource consumption of service requests in cloud environments," the 20th Conference on Innovations in Clouds, Internet and Networks (ICIN), Paris, pp. 169-176, 2017.

[9] Z. Zhang, B. Bockelman, D. W. Carder and T. Tannenbaum, "Lark: Bringing Network Awareness to High Throughput Computing," 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Shenzhen, pp. 382-391, 2015. doi: 10.1109/CCGrid.2015.116

[10] S. Yashpal, S. Alok, " Neural Network in Data Mining", Journal of Theoretical and Applied Information Technology, 37-42, 2009.

[11] C. Krieger, "Neural Networks in Data Mining", technician report, 1996.

[12] S. Duggal, R. Chhabra , "Learning Systems and Their Applications: Future of Strategic Expert System". Issues in Information Systems, Vol. III, 2002.

[13] G. jha, "Artificial Neural Networks," International journal of computer science and issues, Indian Research Institute, PUSA, New Delhi, 2005.

[14] S. Nissen, "Implementation of a fast artificial neural network library (FANN)," Technical report, Department of Computer Science University of Copenhagen (DIKU), 2003.

[15] D. Shanthi, G. Sahoo and N. Saravanan, "Designing an Artificial Neural Network Model for the Prediction of Thromboembolic Stroke," International Journals of Biometric and Bioinformatics (IJBB), Volume (3), 10-18, 2009.

[16] S. Chen, R. Lu, J. Zhang, "An Efficient Fog-Assisted Unstable Sensor Detection Scheme with Privacy Preserved", arXiv:1711.10190v1 [cs.CR] 28 Nov 2017.

[17] I. Rao and E. Huh, "A probabilistic and adaptive scheduling algorithm using system- generated predictions for inter-grid resource sharing," Journal of Supercomputer, 45, pp: 185-204, 2008.

[18] N. Moustafa, I. Al Ridhawi, and A. Hamza, "An Intelligent Dynamic Replica Selection Model within Grid Systems, " in Proc. 8th IEEE GCC conference on Towards Smart Sustainable Solutions, pp. 1-6, February 2015. doi: 10.1109/IEEEGCC.2015.7060061.

[19] "The Cloud Computing and Distributed Systems (CLOUDS) Laboratory, University of Melbourne," December. 2012, Available: http://www.cloudbus.org/cloudsim/.

[20] S. Mostafi, F. Khan, A. Chakrabarty, D. Y. Suh and M. J. Piran, "An Algorithm for Mapping a Traffic Domain Into a Complex Network: A Social Internet of Things Approach," in IEEE Access, vol. 7, pp. 40925-40940, 2019. doi: 10.1109/ACCESS.2019.2906647.

[21] B. Bhuyan, H. Sarma, N. Sarma, A. Kar and R. Mall, "Quality of Service (QoS) Provisions in Wireless Sensor Networks and Related Challenges", Wireless Sensor Network, Vol. 2 No. 11, pp. 861- 868, 2010. doi: 10.4236/wsn.2010.211104.

[22] W. J. Chu Using Chebyshev's Inequality to Determine Sample Size in Biometric Evaluation of Fingerprint Data, Forgotten Books, 2018.

[23] W. Smith, I. Foster, and V. Taylor, "Predicting application run times with historical information," Journal of Parallel Distributed Computing, pp. 1007–1016, 2004.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 418-423 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Biofortification of the Ryegrass (*Lolium multiflorum* L.) with Chicken Manure Compost in the Central Sierra of Peru

Doris Chirinos-Peinado[1*], Jorge Castro-Bedriñana[1], Percy Lara-Schwartz[2]

[1]*Zootechnic Faculty, Universidad Nacional del Centro del Perú, Huancayo, Perú*

[2]*Agrarian Agency of Jauja, Ministry of Agriculture. Junín, Perú*

ARTICLE INFO

Article history:
Received: 01 December, 2019
Accepted: 18 December, 2019
Online: 25 February, 2020

Keywords:
Composting
Manure chicken
Organic fertilizer
Forage quality
Italian ryegrass

ABSTRACT

In Peru, large volumes of chicken manure are produced, whose compost applied at normal levels improves the ryegrass yield and protein intake, to obtain organic products and contribute to animal, environmental and human health. To evaluate the biofortifying effect of Chicken Manure Compost (CMC) incorporated in (T1): 0, (T2): 10 and (T3): 20 t ha^{-1} levels, on Green Forage (GF) yield, Dry Matter (DM), Total Protein (TP) and Plant Height (PH) of Italian ryegrass, in pasture located in Huancayo, Peru (3250 altitude); two cuts were made, the first one was 42 days after the first fertilization and the second was 42 days after the first cut and second fertilization. In second cut, GF yield, in T1, T2 and T3 was 70.05; 61.50 and 97.20 t ha^{-1} (P <0.05); the DM yield was 10.39; 8.44 and 23.34 t ha^{-1}, respectively (P <0.05); TP yield was 1.76; 1.74 and 2.80 t ha^{-1}, respectively (P <0.05) and the PH was 97.0; 93.8 and 99.63 cm, respectively (P <0.05). The application of 10 t ha^{-1} had no greater effect in front of the control. The use of 20 t ha^{-1} of CMC allows to obtain organic forage with higher yield of dry matter and protein and therefore it is possible to obtain organic livestock products, with the advantage of reducing the negative impact of bird manure to public health, when is traditionally eliminated to environment.

1. Introduction

The livestock sector is the world's largest consumer of agricultural land, through grazing and use of forage crops; however, about 20% of pastures have deteriorated due to overgrazing and compaction [1].

Among the most used pastures, ryegrass, in association with alfalfa and clover, are important for feeding cattle and other herbivorous species; in the inter-Andean valleys of Peru they have a useful life of 3 to 5 years, with a biomass production of high nutritional quality [2].

In the Junín region, the area of natural pastures is 1,104 300.04 ha. [3]; 29,649.30 ha are corresponding to cultivated pastures, of which 27,633.20 ha are in the Junín mountain range [3]. The production of ryegrass in 2018 was 196,384 tons and the harvested area was 6,656 ha, with a yield of 29.5ton ha^{-1} [4].

Generally, the improvement of forage production seeks to increase biomass yield and productivity, mainly evaluating the yield of green forage, dry matter and in some cases the plant height; however, it is important to evaluate the nutritional composition, especially the yield and quality of protein. To improve the nutritional contribution, different technological strategies can be used, including agronomic biofortification of basic foods, a process that improves the concentration of one or more nutrients in the edible part of the crop, by applying nutrient-rich fertilizers to the soil or foliage to improve the concentration of micronutrients in the edible part of the crop [5, 6]. Biofortification of ryegrass, through the application of Chicken Manure Compost has enormous potential, as has been reported in the improvement of the contribution of selenium in rice [7, 8].

Italian Ryegrass (*Lolium multiflorum* L.), is a grass of high productive potential, yields more than 10 t ha^{-1} of dry matter of high taste and digestibility, producing up to 6 crops per year, which can be used as green, dehydrated forage or preserved as silage, for which it requires a soil rich in organic matter, especially nitrogen

*Doris Chirinos-Peinado, Av. Mariscal Castilla 3909, El Tambo-Huancayo, +51 924855954 & dchirinos@uncp.edu.pe

[9]; increasing the yield from 4.37 t ha^{-1} to 9.81 t ha^{-1} (5.44 t ha^{-1} more) with application of N_{200} than with N_0 [10].

Traditional fertilization, based on the contribution of nitrogen, phosphorus and potassium, in addition to having a short-term effect [11], affects the physicochemical and microbiological characteristics of the soil; its unbalanced use increases the degradation processes by reducing the organic fraction, causing adverse effects on the flora and fauna of the soil [12]; in addition, its use is increasingly expensive [13] and is not ecologically sustainable, and must be complemented or replaced by organic fertilizers, which improve the nutritional quality of biomass, soil quality and reduce waste contamination chemists [13, 14].

The meat chickens manure constituted by feces, urine, food waste and the material used as a bed, is eliminated causing a negative environmental impact on the air, soil and surface water bodies [15], as an alternative it can be used as compost in soil biofortification, having a series of beneficial effects, the first is the supply of nutrients to plants, improving yield and nutritional quality [16]. The aerobic stabilization of chicken manure, through composting, allows obtaining a quality fertilizer and at the same time contributes to the reduction of environmental pollution [17].

In Peru, the chicken meat production is an important industry; which production is above 150 thousand tons per year, has allowed increasing the per capita consumption of chicken meat from 20 to more than 40 kg per person per year between 2000-2016 [18]. In 2018, the average annual consumption of chicken meat was 47 kg and for 2019 it is expected to be 49 kg [19]; and manure generated by this industry can be used as a nitrogen source and organic matter in agricultural production.

Chicken manure is made up of excretes, food waste, feathers and other bed materials; its N content results from the approximate combination of 70% uric acid and 30% undigested protein [20, 21]. The percentage of nitrogen potentially available in manure from meat chicken, after 7 days of incubation, is 66%, being a very good source of N [22].

The Chicken Manure Compost (CMC), called soil improver, is being used in the fertilization of parks, gardens and ornamental plants; it is obtained from the controlled decomposition of manure from broiler chickens, and can be used for the production of organic forage, improving its quality and soil [23].

Composting is done in aerobic conditions, avoiding the emission of CO_2, ammonia and other volatile compounds into the atmosphere. It has between 20 to 30 % of humidity; 7.0 to 8.5 of pH; 23.8% organic matter; 1.1 % of nitrogen; 1.9% of P_2O_5; 3.4 % of K_2O; 5.1% of CaO; 1.5% of MgO; 0.6 % of Na; 61ppm of boron; 48ppm of Cu; 285ppm of Zn; 565 ppm of Mn y 13 005 ppm of Fe; it is free of pseudomonas, Salmonella and fecal bacteria.

Based on the above and the limited information on the behavior of CMC in the yield and nutritional quality of ryegrass, the objective of this research was to determine the yield of biomass, dry matter and total protein, and the height of Italian ryegrass plant, by incorporating 0, 10 and 20 t ha^{-1} of CMC, to obtain organic and biofortified protein forage, which contributes to obtaining safe animal products that promote animal, environmental and human health.

2. Materials and methods

2.1. Research site and observations

The research was carried out in a six-month-old Italian ryegrass pasture, located in the Yauris experimental farm of the Universidad Nacional del Centro del Perú (UNCP), Huancayo-Junín province (3250 altitude). The experiment lasted 12 weeks, making a first fertilization at the beginning of the experiment and a second fertilization at the sixth week, date on which the second cut of the ryegrass pasture was made, an approximate extension of 1250 m^2, there were located three plots of ryegrass, 160m^2 each.

The first plot was considered as a Control treatment (T1): without fertilization or insemination, the second plot was fertilized with CMC (T2): 10t ha^{-1} and the third one was fertilized with CMC (T3): 20 t ha^{-1}. Two fertilizations were made, the first at the beginning of the experiment, with a first forage cut 42 days after the first fertilization. The second fertilization was made after the first cut and 42 days were expected to make the second cut for the final evaluation of the study parameters. These cutting periods answer to recommendations made for the Mantaro Valley (Junín, Peru), the range of the optimal rest period being 22 to 46 days for cultivated pastures [24].

Figure 1 shown the sequence of the investigation process.

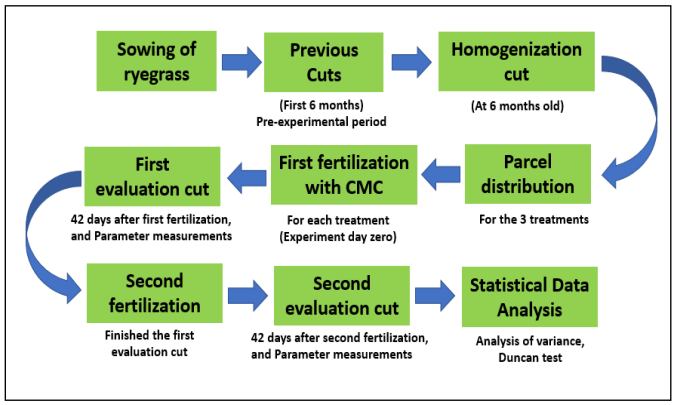


Figure 1: Research procedure secuence

2.2. Sampling and proximal analysis

The study was planned as randomized design with four repetitions, no difference of planting area. To evaluate the yield of green fodder, dry matter and protein in kg m^{-2}, and of the plant height (cm), in each cut, four samples of 1 m^2 $parcela^{-1}$, were taken, being a cutting intensity of 5 cm [25]. Possible weeds were separated, and the weight of the ryegrass was recorded; subsequently, 100 g samples (fresh weight) were taken on an analytical balance and dried in an oven at 65 ° C for 72 hours. The dry matter (DM) and total protein (TP) content of the ryegrass samples was determined in the Animal Nutrition Laboratory of the Faculty of Zootechnics of the UNCP, following the protocols described by the AOAC [26].

2.3 Plant Height (PH)

The PH, before each cut, was measured with a flexometer, for which 10 plants were taken at random per sample per plot. The

cutting intensity was 5 cm from the ground to the last leaf of the stem.

2.4 Statistical analysis

Variance analyzes were performed to determine the statistical differences between the studied variables, under a completely randomized design with three treatments and four repetitions per treatment. The differences between averages per variable were analyzed with Duncan significance tests with a P <0.05. Statistical processing was performed in SPSS v23.

3. Results

The use of 10 and 20 t ha^{-1} of Chicken Manure Compost (CMC) significantly affected the productive parameters of the ryegrass, being the first cut at 42 days after the first fertilization (Table 1 and Figure 2) and the second cut, at 42 days after the second fertilization (Table 2 and Figure 3).

Table 1. Production parameters of Italian ryegrass after 42 days of the first fertilization

Parameters	Treatments	Average	Standard Deviation	95% of the IC for average	
				Lower Limit	Upper Limit
GF yield, kg/m^2	T1: CMC 0 t ha^{-1}	7,600a	1,105	5,842	9,358
	T2: CMC 10 t ha^{-1}	7,475a	0,785	6,226	8,724
	T3: CMC 20 t ha^{-1}	7,575a	1,417	5,320	9,830
	Total	7,550	1,026	6,898	8,202
DM yield kg/m^2	T1: CMC 0 t ha^{-1}	1,223b	0,177	0,940	1,505
	T2: CMC 10 t ha^{-1}	1,183b	0,125	0,984	1,381
	T3: CMC 20 t ha^{-1}	2,153a	0,401	1,515	2,790
	Total	1,519	0,525	1,186	1,853
TP yield, kg/m^2	T1: CMC 0 t ha^{-1}	0,191a	0,028	0,147	0,235
	T2: CMC 10 t ha^{-1}	0,212a	0,022	0,177	0,247
	T3: CMC 20 t ha^{-1}	0,219a	0,041	0,153	0,284
	Total	0,207	0,031	0,188	0,227
Plant Height, cm	T1: CMC 0 t ha^{-1}	76,45b	2,460	72,536	80,364
	T2: CMC 10 t ha^{-1}	83,51a	0,782	82,267	84,758
	T3: CMC 20 t ha^{-1}	84,27a	2,227	80,731	87,819
	Total	81,41	4,088	78,815	84,010

Means per parameter with different letters are statistically different (Duncan, P <0.05)

4. Discussion

4.1 Green Forage (GF) yield

The yield of GF after 42 days of the first fertilization with CMC, does not register significant differences (P >0.05), and sufficient time must pass so that the organic matter and all the nutrients of the fertilizer are progressively processed by the microorganisms of the soil to have a soil improvement effect, as shown in the results obtained after a second application of the CMC in the second cut. The organic phosphorus has to be released by mineralization for plants to absorb it; similarly, the nitrogen that is generally in organic form (> 95%), must evolve to nitrates, to provide nitrogen available for plants in the short term [27, 28].

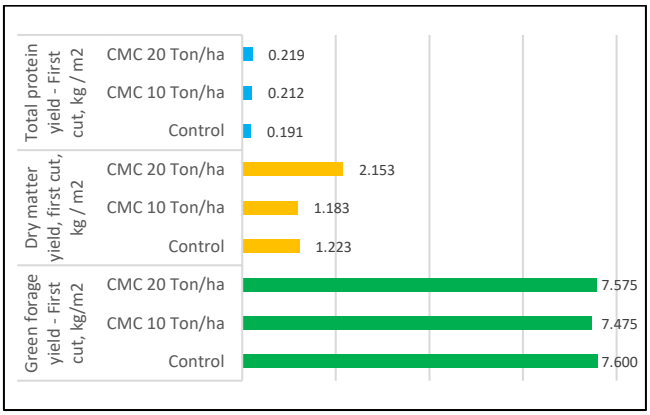


Figure 2: Productive parameters at the first cut, after 42 days of the first fertilization with chicken manure compost (CMC)

Table 2. Production parameters of the Italian ryegrass after 42 days of the second fertilization

Parameters	Treatments	Average	Standard Deviation	95% of the IC for average	
				Lower Limit	Upper Limit
GF yield, kg/m^2	T1: CMC 0 t ha^{-1}	7,005b	0,695	5,899	8,111
	T2: CMC 10 t ha^{-1}	6,150b	0,628	5,150	7,150
	T3: CMC 20 t ha^{-1}	9,720a	1,054	8,042	11,397
	Total	7,625	1,752	6,512	8,738
DM yield kg/m^2	T1: CMC 0 t ha^{-1}	1,039b	0,103	0,876	1,203
	T2: CMC 10 t ha^{-1}	0,844b	0,086	0,706	0,981
	T3: CMC 20 t ha^{-1}	2,334a	0,253	1,931	2,737
	Total	1,406	0,707	0,957	1,855
TP yield, kg/m^2	T1: CMC 0 t ha^{-1}	0,176b	0,017	0,149	0,203
	T2: CMC 10 t ha^{-1}	0,174b	0,018	0,146	0,203
	T3: CMC 20 t ha^{-1}	0,280a	0,030	0,233	0,328
	Total	0,210	0,555	0,175	0,245
Plant height, cm	T1: CMC 0 t ha^{-1}	97,05ab	2,093	93,718	100,381
	T2: CMC 10 t ha^{-1}	93,82b	3,897	87,622	100,027
	T3: CMC 20 t ha^{-1}	99.63a	3,901	93,422	105,838

Total	86.83	3.954	94,322	99,348

Means per parameter with different letters are statistically different (Duncan, p<0.05)

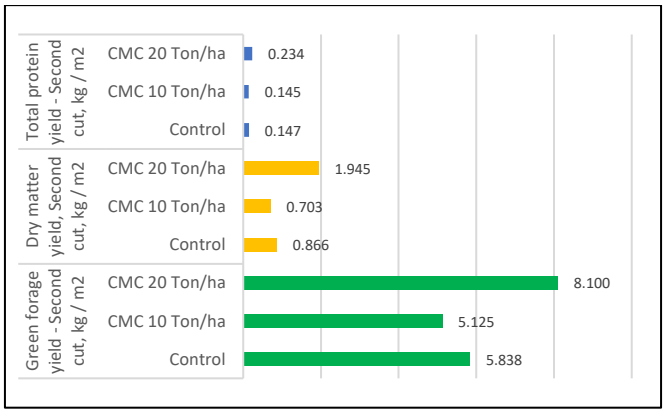


Figure 3: Productive parameters at the second cut, after 42 days of the second fertilization with CMC

The yield of GF m^{-2}, by treatment, after 42 days after the first cut (lapsed 84 days), shows significant differences (P<0.05) due to the use of 20 t ha^{-1} of CMC, increasing from 7.00 kg m^{-2} in plots without fertilization at 8.72 kg m^{-2} in which they received 20 t ha^{-1} CMC. This result shows that CMC, due to its content of organic matter, nitrogen and other nutrients, fertilizes the soil, improves the growth rate; result that agrees with reports where the green forage yield was improved by the use of compost based on municipal solid waste [29], bovine manure [30], organic and inorganic fertilizer [27]; the content of organic matter is increased [30], with greater phosphorus mineralization [31]. Yield of 9.9 and 10.9 t ha^{-1} are reported in treatments with high compost application rates of 33.6 and 44.8 t ha^{-1}, respectively [32], increasing the yield from 4.37 t ha^{-1} to 9.81 t ha^{-1} by the application of N_{200} compared to the control without fertilization [10].

4.2 Dry Matter Yield

The percentage of dry matter of ryegrass, 42 days after starting the experiment (first cut), was between 15.41 and 16.06%, values that differ from other similar investigations [33], which report 37.5% dry matter in ryegrass. Probably due to the agroecological conditions of permanent soil moisture of the Yauris Experimental Farm, which allows the grass to have a higher water content. The DM yield was increased by 1.76 times by the use of 20 t ha^{-1} of CMC in front of the control (P<0.05), result of practical and economic importance, since it allows to increase the animal load and obtain greater productivity of grass.

At the second cut, the dry matter content of ryegrass was between 13.30 and 14.84%, slightly lower values than the one recorded in the first cut. The yield of DM in kg m^{-2} was 2.25 times more when incorporating 20 t ha^{-1} of CMC compared to the control without fertilization. These positive results using compost are reported by applying compost and residual effluents treated in an integrated manner, improving the production of DM from ryegrass [34]. Applying wastewater and olive pomace compost as soil amendments on the growth of the ryegrass, growth parameters were improved by 18.2 and 41.1% at the first and second year of production. A significant increase in total organic carbon, extracted and humidified, and in the humidification, parameters was determined. Ryegrass efficiently used the N content of the amendment and as a result significantly improved dry matter yield [35]. This proper handling of ryegrass fertilization allows obtaining 20 t ha^{-1} of DM [36], similar production obtained here by the use of 20 t ha^{-1} of CMC.

The increase in the yield of DM of ryegrass by CMC effect is of a lot of practical impact, because the contribution of nutrients (organic matter, protein, micronutrients) improves not only the consumption of food, but its digestibility and consequently the animal response; improves soil structure and better soil moisture retention [37]. Broiler chicken manure having a high content of nitrogen product, from high protein diets, between 22.5 to 19.5% [33] not only greatly improve the organic nitrogen content of the soil but also the contribution of P, K, Ca, Mg, Na, B, Cu, Zn, Mn, Fe, among other nutrients, increasing their availability for plants. Organic fertilizers provide between 2 and 4 g of phosphorus per kilogram, enough quantity for forage cultivation [38].

4.3 Total Protein (TP) Yield

The TP content, on a dry basis, at the first cut was between 15.6 to 18.4%; equivalent to fresh content from 2.51 to 2.88%. At the second cut, the percentage of PT on dry basis was between 16.6 and 20.4%, equivalent to fresh values of 2.46 to 2.71%; lower values reported in other experiments [30]; who report that ryegrass has an average of 3.11% of PT on a fresh basis. With these values the protein yield in kg m^{-2} was determined.

At 42 days (first cut), the TP yield was increased by 28 g m^{-2} (P> 0.05) when incorporating the CMC in 20 t ha^{-1}, in relation to the plot that did not receive fertilization. At the second cut, made 42 days after the first cut and second fertilization, when incorporating 20 t ha^{-1} of CMC, the TP yield was increased by 104 g m^{-2} (P< 0.05), with no effect for use of 10 t ha^{-1} of CMC. The increase in the yield of TP due to the use of 20 t ha^{-1} of CMC is of great nutritional and economic importance, because if we scale the results to one hectare, the ryegrass without fertilization, which would produce 1760 kg of PT ha^{-1}, would rise to 2800 kg of PT ha^{-1} per cut, by incorporating 20 t ha^{-1} of CMC; it would have 1040 kg more of TP, which would allow greater supportability of the meadow and animal production.

These results demonstrate that the CMC would have a high potential for nitrogen available for crop uptake [39], being a good alternative for organic forage production and consequently for healthy livestock production, replacing the use of inorganic chemical fertilizers, contributing to the reduction of the negative environmental impact that it would have if the chicken manure is eliminated as garbage [17].

An additional aspect to consider is that in addition to the increase in the content and yield of TP of the ryegrass by the use of 20 t ha^{-1} of CMC, the content of organic matter is also improved, which rises up to 3% by use of bovine manure [30, 27] satisfying the nitrogen requirement (N) of the crops and providing other nutrients, such as phosphorus (P), potassium (K), calcium (Ca) and magnesium (Mg), reported in amendments made with animal waste [28, 40], improves the physical, chemical and biological characteristics of the soil, giving greater stability to the system [41].

4.4 Plant Height (PH)

At 42 days after the first application of the CMC, the PH was significantly improved by the use of 10 and 20 t ha^{-1} compost ($P<0.05$). The PH increased by 7.06 and 7.82 cm due to increased use of the CMC. At the second cut, 42 days after the second fertilization, the PH was improved only in the treatment that considered the incorporation of 20 t ha^{-1} of CMC ($P<0.05$), with no effect by the use of 10 t ha^{-1} of CMC; but the nutritional value was improved. The PH has a certain relationship with the dry matter yield [42].

The Plant Height registered in the present study was similar when different levels of N P K (kg ha^{-1}) were used [43], the values were between 86 and 99 cm in fertilized pastures; while in the control treatment, without fertilization, the PH was only 62 cm; at the Santa Ana-Huancayo Experimental Station (3316 altitude).

These results are of practical importance in animal feeding, since the PH is an indirect indicator of the condition of the pastures; the vigor of a grass can be evaluated by the PH and in the leaves, there is the highest digestible protein content for cattle.

The positive effect about the yield of biomass and protein of ryegrass by the amendment with CMC, is a good alternative for livestock producers, contributing additionally with the recycling of chicken manure and with the carbon capture of the soil [44, 45], eliminating the use of nitrogen fertilizer and reducing the ecological impacts that endanger soil and air pollution [46, 47]; ryegrass cultivation being a good forage option, due to its high yield, nutritional quality and ability to grow in a great diversity of soils [17].

5. Conclusion

The yield of green forage, dry matter, total protein and ryegrass height were significantly improved ($P<0.05$) when using 20 t ha^{-1} of chicken manure compost, being possible to replace inorganic fertilizers and obtain an organic forages production, reducing the negative impact by traditionally eliminating bird manure, in addition to improving soil condition.

Acknowledgments

The authors thank Dr Masataka Shimojo, Kyushu University, for experimental advice, Hermenegildo Dorregaray and technical staff of the Agricultural Farm of Yauris of the Universidad Nacional del Centro del Perú, for support during the experiment.

Conflict of interest

All the authors declare that there is no conflict of interests regarding the publication of this manuscript.

References

[1] FAO. La ganadería y el medio ambiente. FAO ed. Roma-Italia. 2019. http://www.fao.org/livestock-environment/es/.

[2] Aliaga, L., Moncayo, R., Rico, E. & Caycedo, A. Producción de cuyes. Lima: Fondo Editorial de la Universidad Católica Sedes Sapientiae. Lima. Perú. 2009.808 p.

[3] DEIA. Compendio Estadístico Agropecuario de Junín 1,997 – 2,016. Dirección de Estadística e Información Agraria-Dirección General de Agricultura de Junín. 2017. https://drive.google.com/file/d/0B6teYThL_oDqX2hBazIwVmlZbFk/view.

[4] SIEA, 2019. Sistema Integrado de Estadística Agraria. Datos en Excel Anuario de Producción Agrícola 2018. Ministerio de Agricultura. http://siea.minagri.gob.pe/siea/?q=noticias/anuario-de-produccion-agricola-2018

[5] Saeid, A., & Jastrzębska, M. Agronomic biofortification as a key to plant/cereal fortification in micronutrients. In: Food Biofortification Technologies [Internet]. CRC Press; 2017 [cited 2019 Jan 5]. p. 1–60. https://www.taylorfrancis.com/books/9781498756600/chapters/10.1201/9781351228350-1.

[6] Lidon, F. C., Oliveira, K., Ribeiro, M. M., Pelica, J., Pataco, I., Ramalho, J. C., et al. "Selenium biofortification of rice grains and implications on macronutrients quality". J Cereal Sci 81:22-9, 2018. DOI: https://doi.org/10.1016/j.jcs.2018.03.010.

[7] White, P. J., & Broadley, M. R. "Biofortification of crops with seven mineral elements often lacking in human diets - iron, zinc, copper, calcium, magnesium, selenium and iodine". New Phytologist, 182: 49–84, 2009. https://doi.org/10.1111/j.1469-8137.2008.02738.x.

[8] Wang, Y. D., Wang, X., & Wong, Y. S. Generation of selenium-enriched rice with enhanced grain yield, selenium content and bioavailability through fertilization with selenite. Chin. Univ. Food Chem. 141, 2385-2393, 2013. https://doi.org/10.1016/j. foodchem.2013.05.095.

[9] Kramberger, B., Gselman, A., Podvrsnik, M., Kristl, J., & Lesnik, M. "Environmental advantages of binary mixtures of Trifolium incarnatum and Lolium multiflorum over individual pure stands". Plant, Soil and Enviroment vol. 59:22-28, 2013

[10] Rechițean, D., M. Dragoș, N. Dragomir, M. Horablaga, M. Sauer, D. Camen, I. Toth, & A. Sala. "Associated Culture of Italian Ryegrass (Lolium multiflorum) and Crimson Clover (Trifolium incarnatum) under Nitrogen Fertilization". Animal Science and Biotechnologies. 51(1):129-132, 2018.

[11] Cueto, W. J., Quiroga, G. H., & Becerra, M. C. "Nitrógeno disponible y desarrollo del ballico anual. 1. Producción, calidad del forraje y acumulación de nitratos". Terra. 21(2): 285-295, 2003.

[12] Brack, A., & Mendiola C. La contaminación ambiental, Enciclopedia Virtual "Ecología del Perú". Disponible en: http://www.peruecologico.com.pe/libro.htm.Tomado el 20/ene/2019.

[13] Cortijo, E., Faure, G., & Le Gal. P. Inserción de las pequeñas explotaciones familiares en la cadena de suministro de los lácteos en el Valle del Mantaro (Perú): hacia una gestión de apoyo que tome en cuenta la diversidad de los actores. Cirad. La Recherche Agronomique Pour le Développement. 2010. http://siteresources.worldbank.org/EXTSOCIAL DEVELOPMENT/Resources/244362-1265299949041/6766328-1308081363318/Dairy-Study-Final-SPA-ENG.pdf.

[14] Gordillo, F., Peralta, E., Chavez, E., Contreras, V., Campuzano, A., & Ruiz, O. "Producción y evaluación del proceso de compostaje a partir de desechos agroindustriales de *Saccharum officinarum* (caña de azucar)". RIA. 37(2):140-149, 2011. https://www.redalyc.org/articulo.oa?id=86421189006.

[15] Pinos-Rodríguez, J. M., García-López, J. C., Peña-Avelino, L. Y., Rendón-Huerta, J. A., González-González, C., & Tristán-Patiño, F. (2012). Impactos y regulaciones ambientales del estiércol generado por los sistemas ganaderos de algunos países de América. Agrociencia 46: 359-370.

[16] Kupper, T., Bürge, D., Bachmann, H. J., Güsewell, S., & Mayer, J. Heavy metals in source-separated compost and digestates. Waste Management, 34(5), 867–874, 2014. https://doi.org/10.1016/j.wasman.2014.02.007

[17] Riera, N. I., Della Torre, V., Rizzo, P. F., Butti, M., Bressan, F. M., Zárate, N., Weigandt, C. & Crespo, D. "Evaluación del proceso de compostaje de dos mezclas de residuos avícolas". Rev FCA UNCUYO 46(1):195–203, 2014. http://revista.fca.uncu.edu.ar/images/ stories/pdfs/2014-01/Cp14_Riera.pdf.

[18] Bustamante, C., Gutiérrez, E., Malla, J., Palacios, R., & Revilla, F. Planeamiento Estratégico para el Subsector del Pollo de Engorde en el Perú. Tesis Magíster Administración de Negocios Globales. 2018. EPG-PUCP. http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/11641/BUSTAMANTE_GUTIERREZ_PLANEAMIENTO_POLLO.pdf?sequence=1.

[19] aviNews. Avicultura de Perú continúa creciendo este año 2019. La Revista Global de Avicultura. https://avicultura.info/avicultura-de-peru-continua-creciendo-este-ano-2019/

[20] Huff, W. E., Malone, G. W., & Chaloupka, G. W. "Effect of litter treatment on broiler performance and certain litter quality parameters". Poult. Sci. 63:2167–2171, 1984.

[21] Nahm, K. H. "Evaluation of the nitrogen content in poultry manure". Worlds Poult. Sci. J. 59:77–88, 2003.

[22] Ruiz Diaz, D. A., Sawyer, J. E., & Mallarino, A. P. "Poultry Manure Supply of Potentially Available Nitrogen with Soil Incubation". Agronomy Journal, 100(5):1310, 2008. doi:10.2134/agronj2007.0371.
[23] San Fernando. Compost de Guano de Pollo, Compuesto Natural. Mejorador de suelos. San Fernando S.A. Lima. Perú. 2007.
[24] Mamani, E., Jiménez, A. R., San Martín, H. F., Huamán, U. H., Ara, G. M., Carcelén, C., & Huamán, C. A. "Determinación del periodo óptimo de descanso de la pastura asociada Lolium multiflorum, Trifulium pratense y Medicago sativa, pastoreada por cuyes en la sierra central del Perú". Revista de Investigaciones Veterinarias del Perú, 26(3):404-411, 2015. https://dx.doi.org/10.15381/rivep.v26i3.11174.
[25] Ferret, A. "Control de calidad de forrajes". En: XIX curso de especialización FEDNA. Madrid. Fundación Española para el desarrollo de la nutrición animal. 2003. http://fundacionfedna.org/sites/default/files/03CAP_VII.pdf.
[26] AOAC. Official Methods of Analysis of the Association of Official Analytical Chemists. Vol 1. 15th ed. Washington DC: AOAC. 1990. 1298 p.
[27] Hernández, M. T., Chocano, C., Melgares, J., González, D., & García, C. "Incidencia de enmiendas orgánicas sobre la calidad del suelo en ciruelo ecológico". VII Congreso SEAE. Zaragoza, España N°26, 2006. https://docplayer.es/72880188-Incidencia-de-enmiendas-organicas-sobre-la-calidad-del-suelo-en-ciruelo-ecologico.html.
[28] Aguirre, M. E., Bussetti, S. G., & Santa María, R. M. "Fósforo hidrosoluble y disponible en el suelo por el agregado de cáscara de girasol". Rev. Inv. Fac. Cienc. Agra. 11:1-6. 2007.
[29] Herencia, J. F., Ruiz, P. J., Melero, P. S., García Galvis, A., Morillo, E., & Maqueda, C. "Comparison between organic and mineral fertilization for soil fertility levels, crop macronutrient concentrations and yield". Agron J. 99: 973-983, 2007.
[30] Salazar, S. E., Trejo, E. H., López, M. J., Vázquez, V. C., Serrato, C. J., Orona, C. I., & Flores, M. J. "Efecto residual de estiércol bovino sobre el rendimiento de maíz forrajero y propiedades del suelo". Terra Latinoamericana. 28: 381-390, 2010.
[31] Jiménez, L., Larreal, M., & Noguera, N. "Efectos del estiércol de bovino sobre algunas propiedades químicas de un Untisol degradado en el área de la Machiques Colón, estado de Zulia". Rev. Fac. Agron. (Luz)21:1-11, 2004.
[32] Smiciklas, K. D., Walker, P. M., & Kelley, T. R. "Evaluation of compost for use as a soil amendment in corn and soybean production". Compost Science and Utilization 16(3):183-191, 2008.
[33] Castro, J., & Chirinos, D. Manual de Formulación de Raciones Balanceadas Para Animales. 2008. Concytec. 230 p.
[34] Chipula, G., Sakrabani, R., Hess, T., & Tyrrel, S. "Compost-Sewage Effluent Integration for Ryegrass Production. Journal of Crop Improvement" 28(3):345-360, 2014. DOI:10.1080/15427528.2014.888532).
[35] Montemurro, F., Convertini, G., & Ferri, D. "Mill wastewater and olive pomace compost as amendments for rye-grass". Agronomie, EDP Sciences; 24 (8):481-486, 2004.
[36] Donaghy, D., & Fulkerson, B. Principles for developing an effective grazing management system for ryegrass-based pastures. Tasmanian Institute of Agricultural Research, Burnie, Tasmania. 2001. 10 p.
[37] Moral, R., Paredes, C., Bustamante, M., Marhuenda-Egea, A.F., & Bernal, M.P. "Utilization of manure composts by high-value crops: Safety and environmental challenges". Bioresource Technology 100(22): 5454–5460, 2009.
[38] Eghball, B. "Soil properties as influenced by phosphorus- and nitrogen- based manure and compost applications". Agron. J. 94: 128-135, 2002.
[39] Nahm, K. H. "Factors influencing nitrogen mineralization during poultry litter composting and calculations for available nitrogen". Worlds Poult. Sci. J. 61:238–255, 2005.
[40] Pérez, A., Céspedes, C., & Núñez, P. "Caracterización física-química y biológica de enmiendas orgánicas aplicadas en la producción de cultivos en República Dominicana". R.C. Suelo Nutr. Veg. 8(3):10-29, 2008.
[41] Donn, S., Wheatley, R., McKenzie, B., Loades, K., & Hallett, P. "Improved soil fertility from compost amendment increases root growth and reinforcement of surface soil on slopes". Ecological Engineering, 71:458-465, 2014. http://www.sciencedirect.com/science/article/pii/S0925857414003577.
[42] Villalobos, L., & Sánchez, J. "Evaluación agronómica y nutricional del pasto Rye grass perenne tetraploide (*Lolium perenne*) producido en lecherías de las zonas altas de Costa Rica: Producción de biomasa y fenología". Agronomía Costarricense 34(1): 31-42, 2010. ISSN:0377-9424. http://www.scielo.sa.cr/pdf/ac/v34n1/a03v34n1.pdf.
[43] Nestares, P. A., Camargo, S. M., & Rojas, S. A. Producción de semillas de Rye grass italiano con riego tecnificado. INIA. EEA Santa Ana. Huancayo, 1992. https://alicia.concytec.gob.pe/vufind/Record/INIA_73365a37167df77a238e3ebeb2c0eb02.
[44] Fowler, C. J., Condron, L. M., & McLenaghen, R. D. "Effect of green manures on nitrogen loss and availability in organic cropping system". New Zealand J. Agri. Res. 47, 95-100, 2004
[45] Zhang, M., Gavalak, R., Mitchell, A., & Sparrow, S. "Solid and liquid cattle manure application in a subartic soil". Agron. J. 98, 1551-1558, 2006.
[46] Zhang, H., Xu, M., & Zhang, F. "Long-term effects of manure application on grain yield under different cropping system and ecological conditions in China". J. Agric. Sci. 2009, 147, 31-42, 2007.
[47] Jacinthe, P. A., & Lal, R. "Carbon storage and minesoil properties in relation to topsoil application techniques". Soil Sci. Soc. Am. J. 71, 1788-1795; 2007.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 424-430 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Fuzzy Logic Control Management with Stand Alone Photovoltaic - Fuel Cell System

Ehab Mohamed Zakaria[1,*], Shawky Hamed Arafa[1], Maged Naguib Fahmy Nashed[2], Salah Ghazi Ramadan[1]

[1]*Faculty of Engineering, Benha University, Egypt*

[2]*Electronics Research Institute, Cairo, Egypt*

ARTICLE INFO

Article history:
Received: 23 December, 2019
Accepted: 12 February, 2020
Online: 25 February, 2020

Keywords:
Hybrid system
Photovoltaic
Fuel cell
Electrolyzer
FLC

ABSTRACT

In this paper, a Hybrid System introduced (HS), which consists of Solar Photo-Voltaic (PV) system and the Fuel Cell (FC) System as alternative source of electrical power. The HS is used to supply the required electricity to a Single-Phase load with the help of Pulse Width Modulation Voltage Source Inverter (PWM-VSI). The PV provide the FC system with the required power to be used when it is needed to feed the loads during the absent of the solar sun lights. The FC system generate the Hydrogen (H_2) from electrolyzer which Supplied with the Photo-Voltaic energy and water (H_2O). While, the generated H_2 is reserved in a Tank which provides the FC Stack by the required Hydrogen level. The output voltage from the FC Stack then, Boosted and converted to AC voltage suitable for the loads using a PWM-VSI. The fuzzy logic control (FLC) is designed to control management between level of tank H_2 and power demand of load. The individual components are simulated in MATLAB/Simulink environment and tested, also, complete system modeling is presented, and the results obtained are found to be satisfactory.

1. Introduction

A Standalone Hybrid System (HS) supplies the load by electricity without connection to the centralized power stations. Hybrid systems (HS) used in remote areas, where there are no electricity sources available therefore, selection of hybrid systems (HS) are Suitable for these areas. The paper purpose is to simulate and model the various components of a Photovoltaic Fuel-Cell (PVFC) hybrid system which is preferred to provide the demands of electric for areas far from the grid. Modeling and simulations for the presented system are performed using Simulink/ MATLAB package and results are presented to verify the effectiveness of this system.

The photo-voltaic energy (PV) is widely used in low power system applications among the renewable energy sources. PV generators have many advantages such as being pollution free, long life, silent, time etc. photo-voltaic (PV) have variable output power depending on the weather conditions variations [1, 2]. One way for overcoming this problem is to integrate and Combine the photo-voltaic array with other power sources for example, Fuel cell (FC). The selection of FC beak up power supply is a very attractive when it used with an intermittent source of power generation such as photo-voltaic power due to several attractive features of FC like fast load response and efficiency [3, 4]. PV installed considered as one of the alternative energy sources for clean generated energy. The power produced in a PV is in DC power form. Whereas certain cases may allow the direct using of the generated DC power from PV, while frequently required to convert the generated DC power to an AC power for control and integration into load system.

A Hybrid Systems consisting of PV-wind-diesel is used in [5, 6]. But due to increasing cost of diesel and use of the diesel environmental pollution is created that makes the operation of diesel engine prohibited. Another type of Hybrid Systems consisting of a wind-hydro is fed grid and inverter used to fed AC load where the power converters are used to improve the precision of the power system [7, 8].

FLC has been employed as a maximum power point tracking tool in [9]. The FLC was also used as a tool for power management in the standalone PV power system described in [10]. The work presented in [11] shows that the stand alone a hybrid power system and can be improved with FLC when it is used as management power for the best operation.

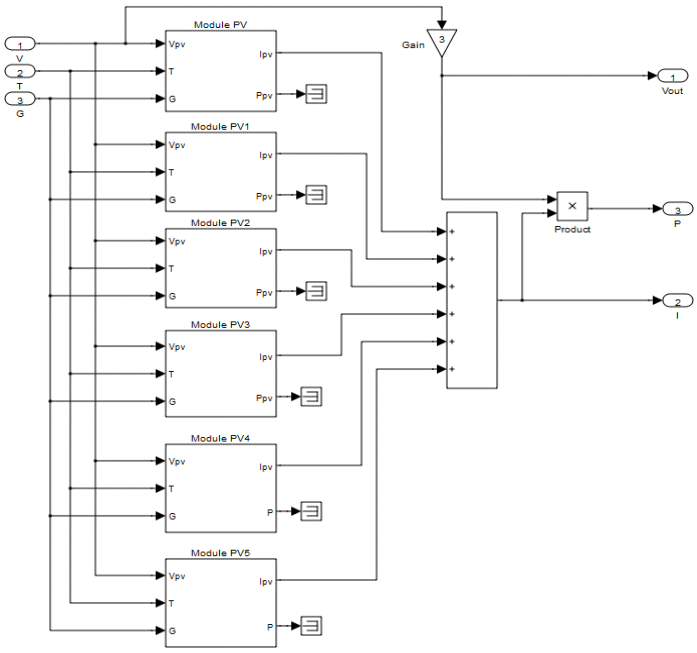
* Ehab Mohamed Zakaria, Email: Eng_ehab_sakr@yahoo.com

2. A Dynamic Model of Hybrid System

The equivalent circuit of the PV and the various equations leading to equation are provided in [1];

$$I = I_{PH} - I_S\left(\exp\left[\frac{q(V+IR_S)}{kTA}\right] - 1\right) - \frac{(V+IR_S)}{R_{SH}} \quad (1)$$

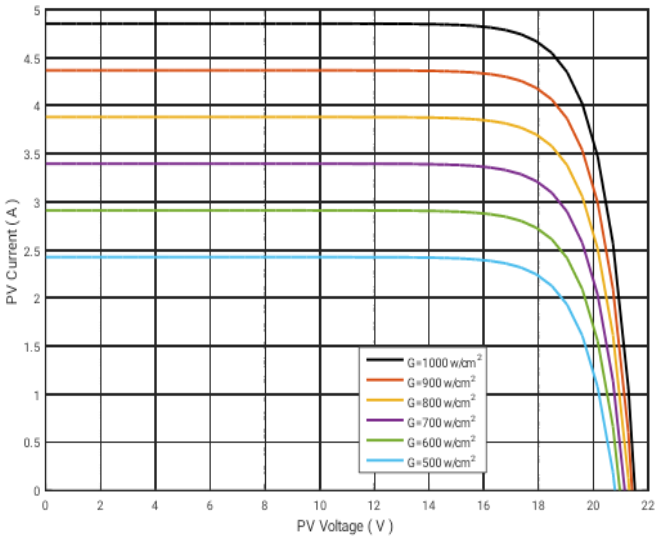
As shown in Figure 1, the Simulink model of the module and PV generator is Consists of diode, photo current source, series and parallel resistor which describing the internal resistance to the current flow and expressing the leakage current. All data of system is written in our paper [11]. In Figure 2, the V-I characteristic curves of the simulated module at various irradiance levels and at various PV cell temperatures. It is seen that at constant PV cell temperature, increasing the irradiance will increase the short circuit current largely and increase the open circuit voltage slightly.

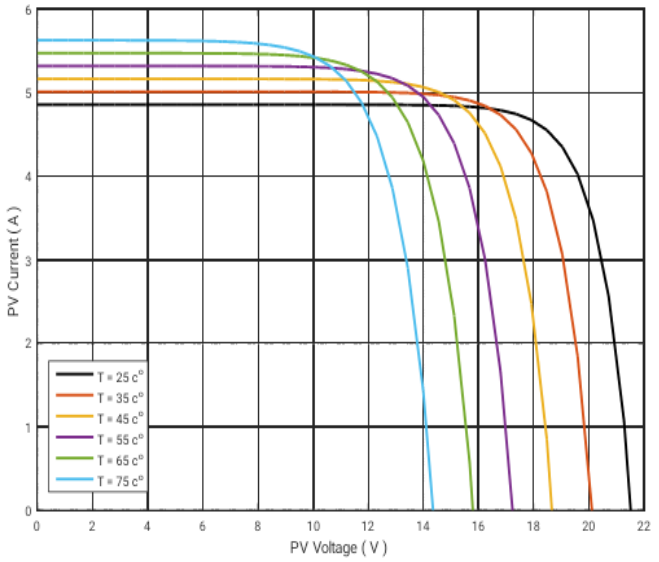


The parts of the system as Proton Exchange Membrane Fuel-Cell (PEMFC), Hydrogen Storage System data and equations are written in reference [11]. Figure 3, is shown the voltage-current density (V-I) characteristic of the FC module. While Figure 4, is shown a complete Simulink model for the Electrolyzer.

Since variable voltage needed so, the buck-boost converter is preferred. It is converting the supply voltage source to higher and lower voltages at the load terminals. The voltage of load terminals is controlled by continuously adjusting the duty ratio of the buck-boost converter power switch. Since the voltage polarity at the load terminal is opposite to the voltage polarity at the source terminal for most DC/DC converter types. In steady state condition, the inverting type buck-boost converter output voltage is given in Equation [2];

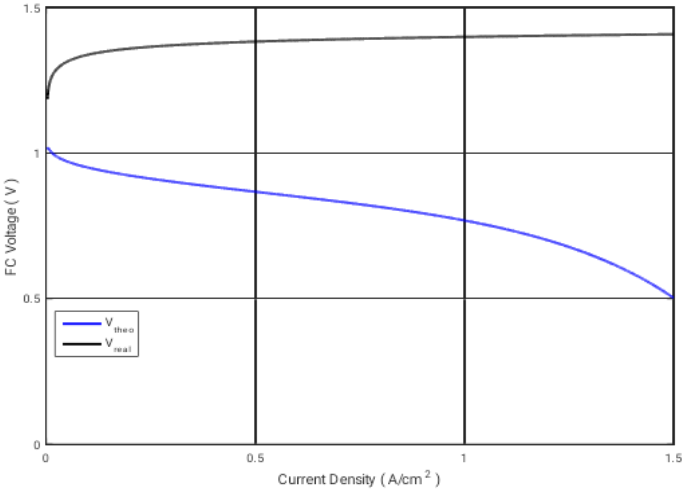
$$\frac{V_o}{V_i} = -\frac{D}{1-D} \quad (2)$$

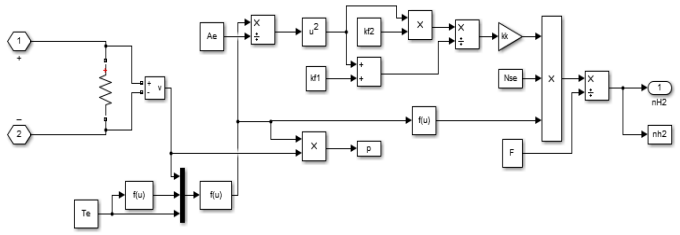




(b) at different temperatures

Figure 2. V–I characteristic curves of the PV module





Thus, the output voltage can be regulated by controlling duty ratio of the used semi-conductor output voltage to higher or lower than the voltage of the source through properly control the MOSFET power switch operating duty ratio. The characteristic of the buck-boost depends on the value of the capacitor and inductor. The considered factors to determine the value of inductance include the input voltage range, the output voltage range, the inductor maximum current, and the switching frequency. The capacitor is used to regulate the voltage. The capacitance depends on the inductor current ripple, the switching frequency, and the desired output voltage ripple. Figure 5 is showing the Buck-Boost converter DC control model by MATLAB/Simulink.

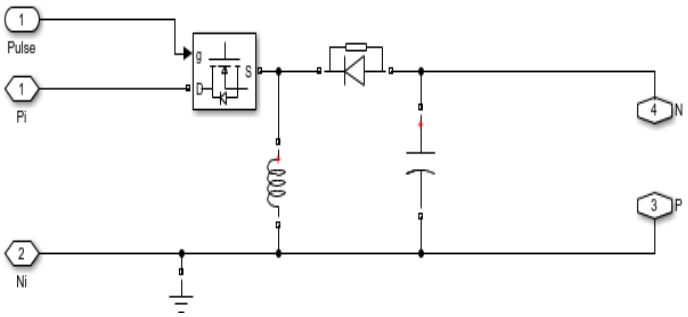


Figure 5. Buck/Boost DC Converter.

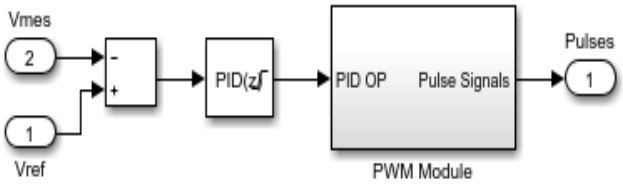
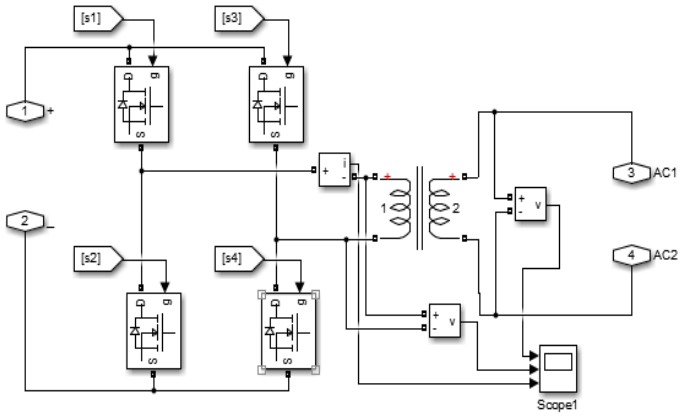


Figure 6. Pulse and Regulation Model



The output voltage regulation of the Buck-Boost Converter is carried out using a PID controller as shown in Figure 6. Then, the controller output used as reference value for the PWM module to generate the required pulses with a suitable duty cycle.

Inverter modelling is a simple topology with a IGBT H-bridge inverter and a linear transformer which is simple to be implemented and completely isolated. The inverter converts the DC/AC with RMS value and Frequency suitable for the connected loads. Figure 7, indicate the MATLAB/Simulink model for the single phase PWM VSI with linear transformer.

3. Control System

3.1. MPPT Algorithm

Figure 8 a present perturb and observe algorithm, (P&O) while a simply changes a variable X, is a variable in current, voltage or duty cycle (D). to obtain the next X, ΔX stay as it during ΔT, then obtain power output of PV array.

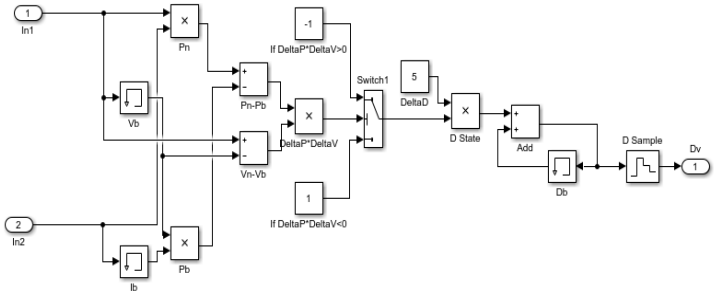


Figure 8. P&O MPPT Algorithm Modules

3.2. FLC Management

The order by FLC offer the advantage of being robust, efficient and works to the control management between the level of tank H_2 and the power demand of the load FL is control of the ON/OFF inverter switch. There are different ways to implement a fuzzy regulator but in general the presentation adopted splits up in three steps: the fuzzification, inference and defuzzification as shown in figure 9. Figure 10, shows the implementation of our fuzzy regulator, [12]-[13].

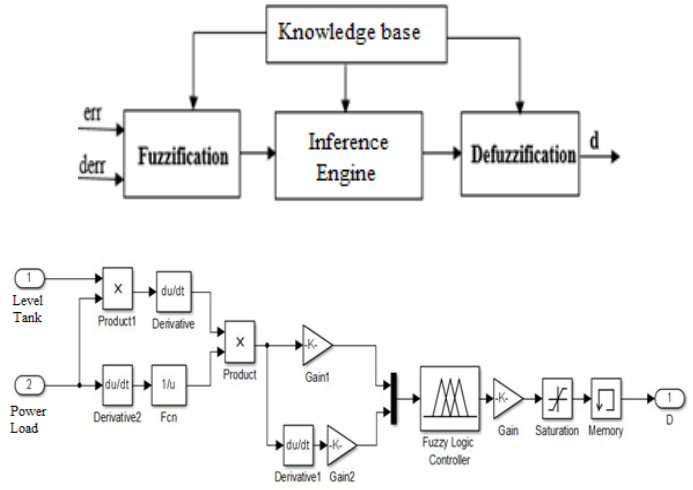


Figure 10. Simulink model of fuzzy control.

4. System and Results

All Simulation done in Simulink/MATLAB by making each component separately then combines them together. All system blocks have been studied and implemented to ensure Each part of the system block is enough precise to run the MATLAB/simulation and giving the adequate results. The PV module array for solar panels, Buck-Boost convertor with MPPT and Regulation, electrolyzer, Hydrogen tank, FC Stack, and the single-phase inverter have been created and well matched with each other. Figure 11, is showing the MATLAB/Simulink model for the whole hybrid system. The PV Array used to provide the FC system with its requirements from the power through the Electrolyzer which generate the required Hydrogen to be used as input for the Fuel Cell Stack to generate its output terminal voltage to feed the loads connected to its terminals via a single phase PWM VSI.

The maximum voltage from PV is designed to be 300 V, nevertheless the PV panel maximum power (MP) is varying due to weather, a (P&Q) MPPT algorithm is used to get the maximum power (MP) from the photo-voltaic panels via the Buck/Boost Converter. In this simulation, the photo-voltaic panels maximum voltage and current were obtained and the electrolyzer became able to generate hydrogen with maximum output. The generated Hydrogen is stored at Storage Tank which is connected to the FC Stack. The Fuel-Cell Stack when fed with the required Hydrogen level it produces its Actual output voltage which depends on the Hydrogen pressure value and its current density. The output voltage from the FCs input to a Buck-Boost DC/DC converter to step-up or down its voltage to a stabilized level suitable for the inverter input terminals. The inverter output AC voltage is used to fed the loads connected its output terminals.

Two scenarios used to examine the system. First, the design performance is observed and when the irradiance and temperature variations with and without MPPT. Secondly, the entire system performance is observed when constant irradiance (1000 W/m^2) and temperature (25^{o}C) and load variable.

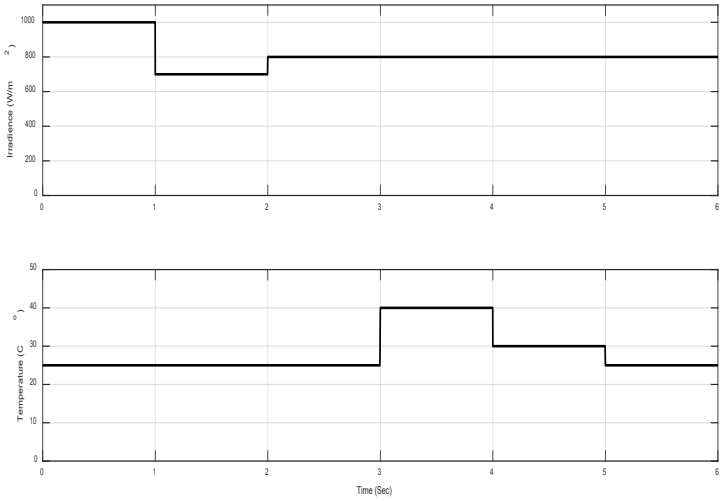


Figure 12. the irradiance and temperature variations with time

4.1. Complete System with Irradiance and Temperature Variations:

The result obtained in Figure 12, suggests that the irradiance and temperature variations.

a) The system response without MPPT

In Figure 13, the PV voltage and Buck-Boost voltage of PV are shown. While Figure 14, is shown FC voltage and converter voltage of FC. Electrolyzer mass flow and tank mass of H_2 are shown in Figure 15. Figure 16, is shown the load current and voltage.

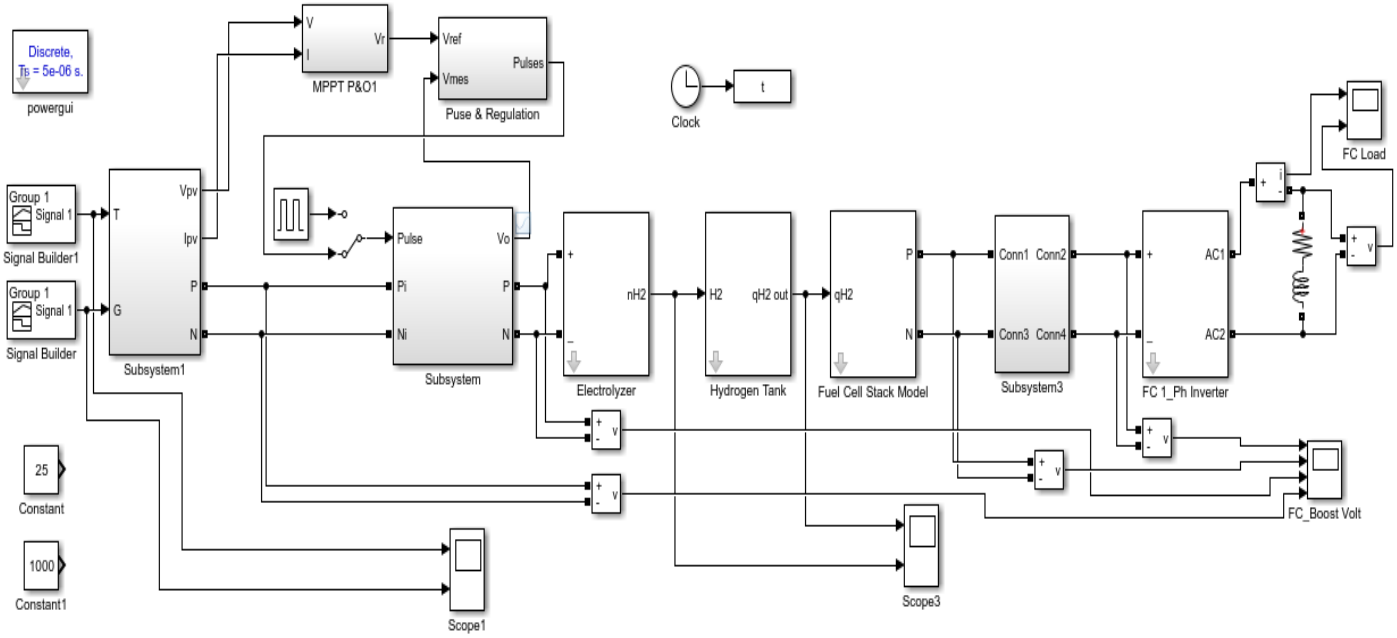


Figure 11. Simulink model for the whole hybrid system

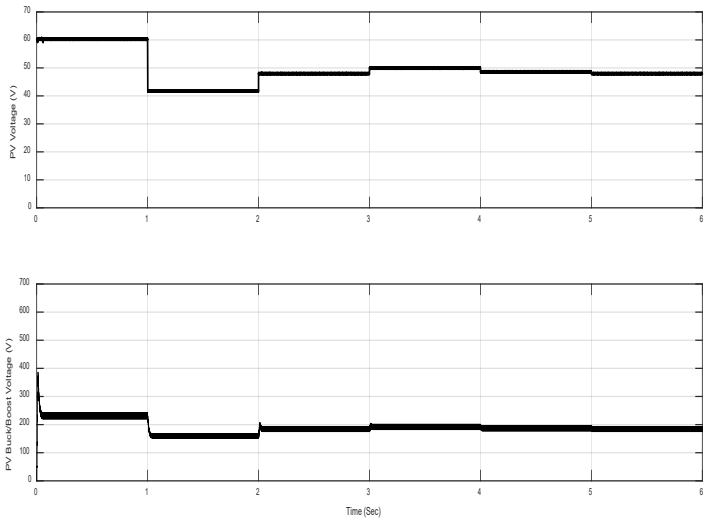


Figure 13. PV voltage and Buck-Boost voltage.

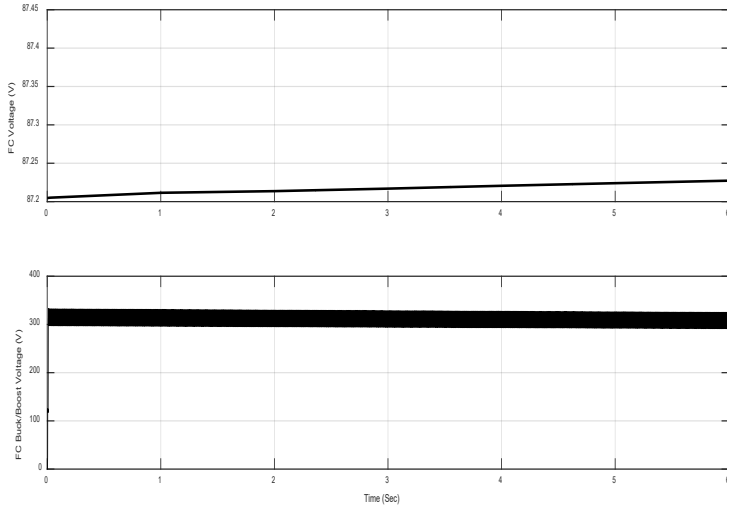


Figure 14. FC voltage and converter voltage of FC.

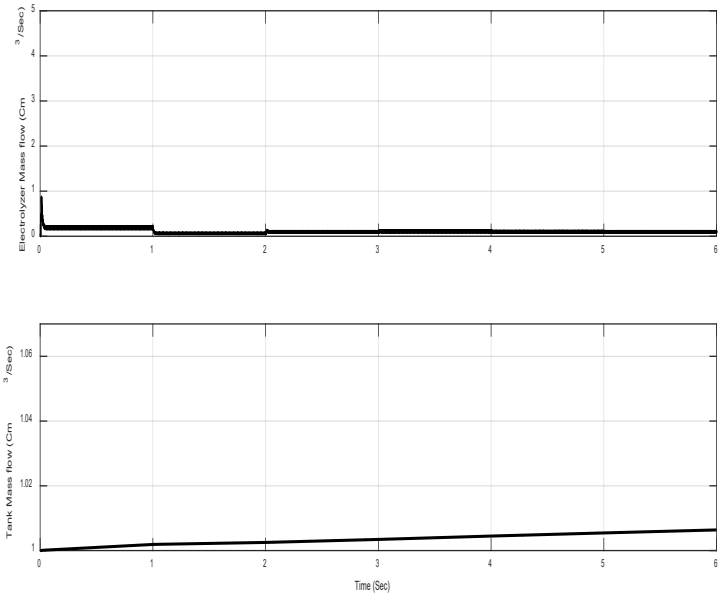
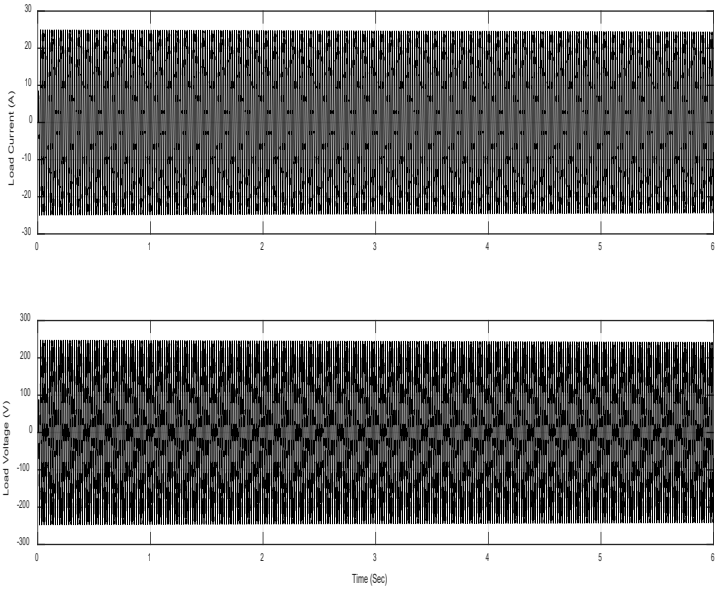


Figure 15. Electrolyzer mass flow and tank mass of H_2.



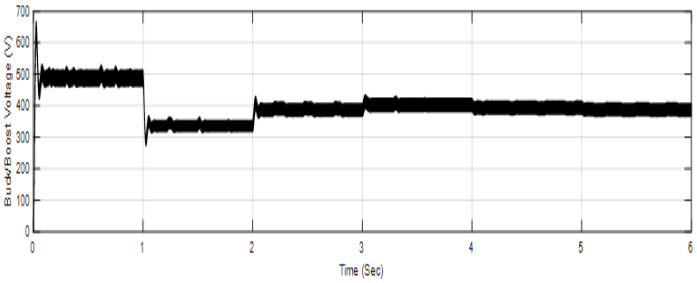


Figure 17. Buck-Boost voltage of PV with MPPT.

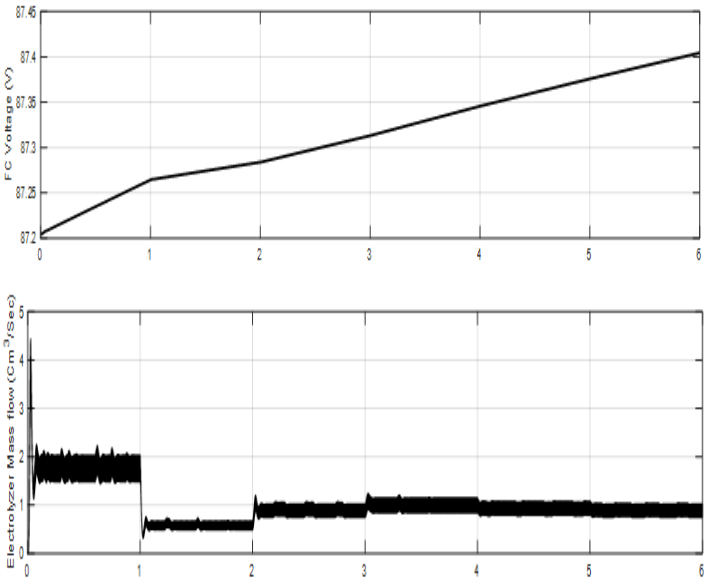


Figure 18. FC voltage and Electrolyzer mass flow with MPPT.

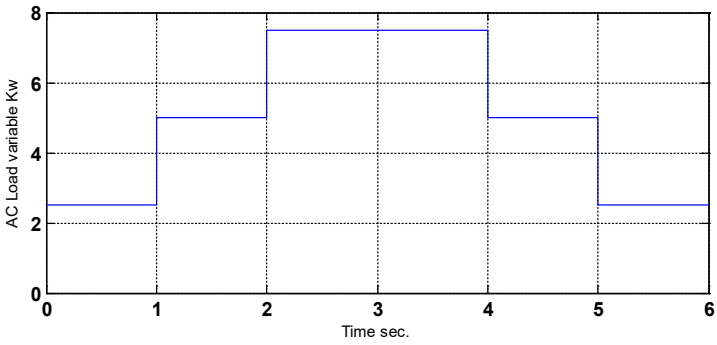


Figure 19. AC load variable with time

Figure 20, is shown the PV and Buck-Boost converter voltage. FC voltage and FC converter voltage are shown in Figure 21. While, Electrolyzer mass flow and tank mass of H_2 are shown in Figure 22. Figure 23, is shown the load current and voltage.

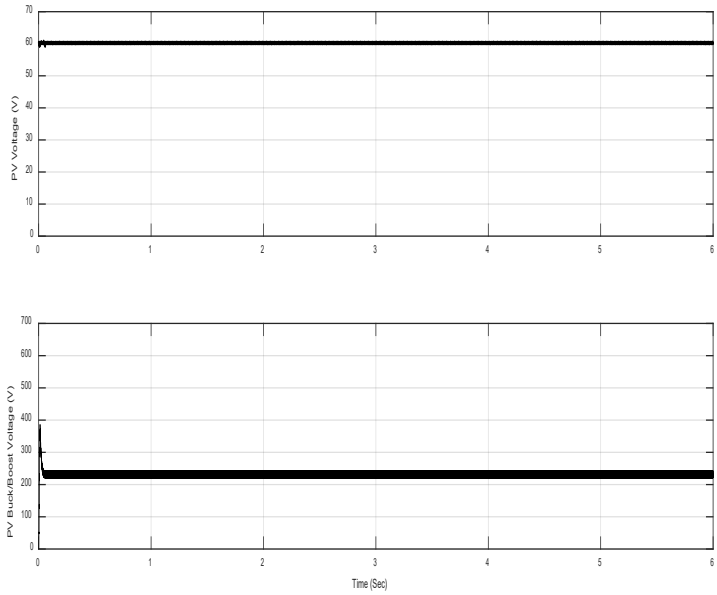


Figure 20. PV voltage and Buck-Boost voltage.

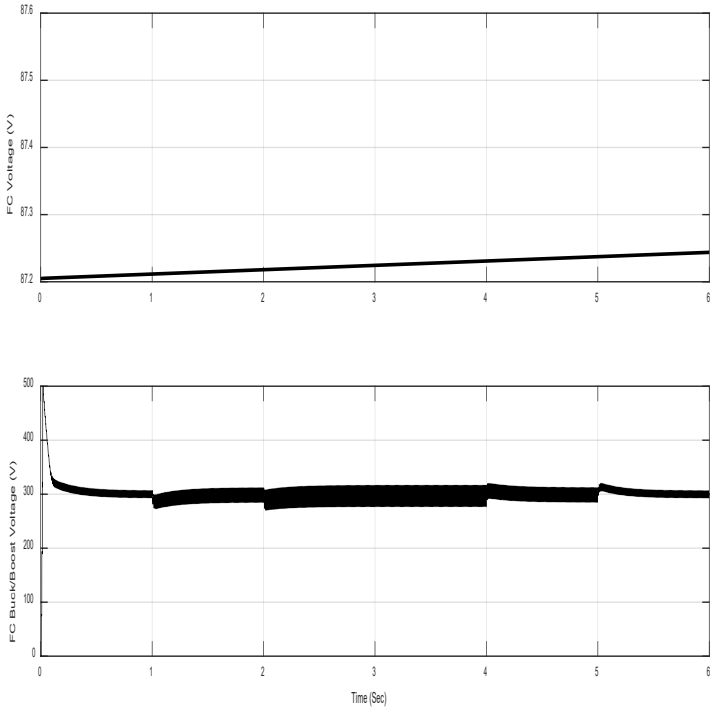


Figure 21. FC voltage and FC converter voltage.

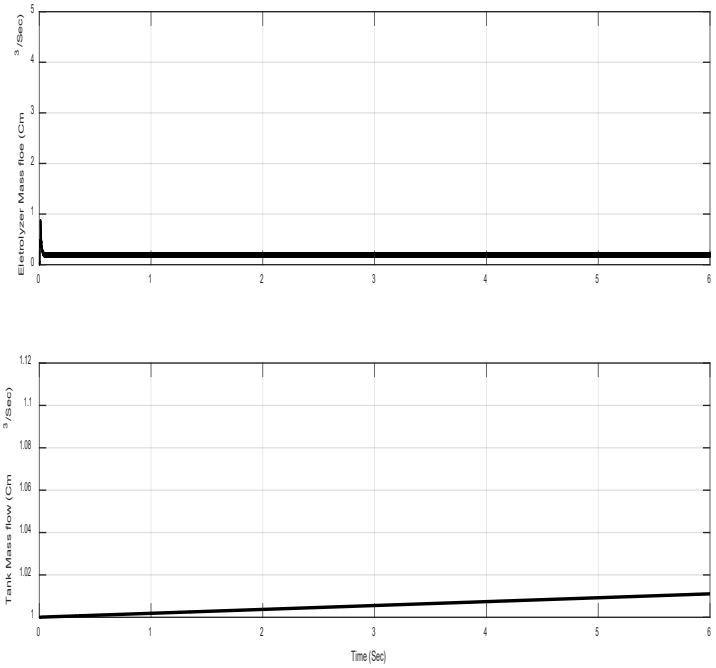


Figure 22. Electrolyzer mass flow and tank mass of H_2.

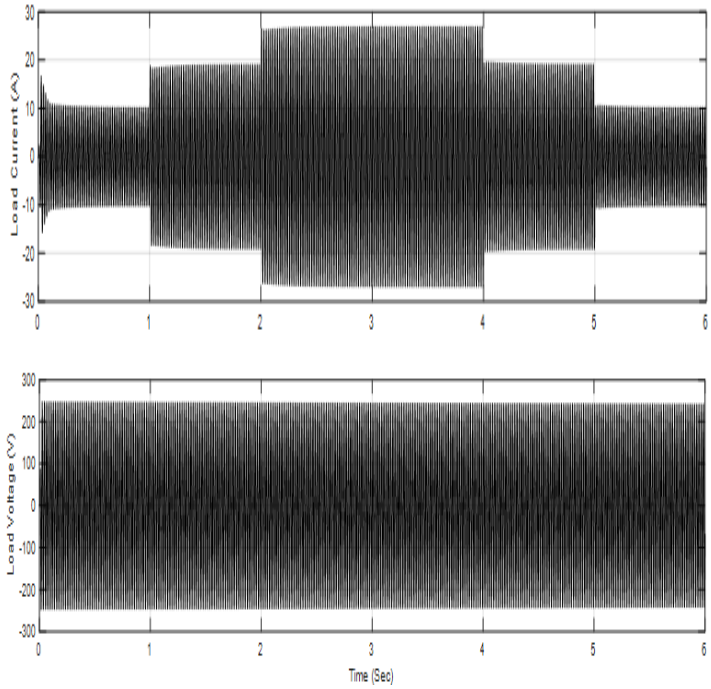


Figure 23. load current and voltage.

b) The system response with MPPT

In this case, the Buck-Boost voltage of PV and FC voltage are shown Figure 24, Buck-Boost voltage is increase to the value of MPPT. Figure 25, is shown the Electrolyzer mass flow and tank mass flow.

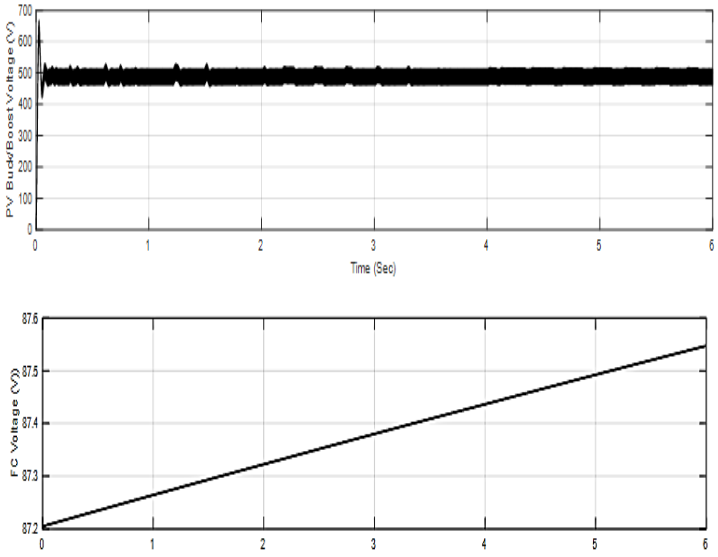


Figure 24. Buck-Boost voltage of PV and FC voltage.

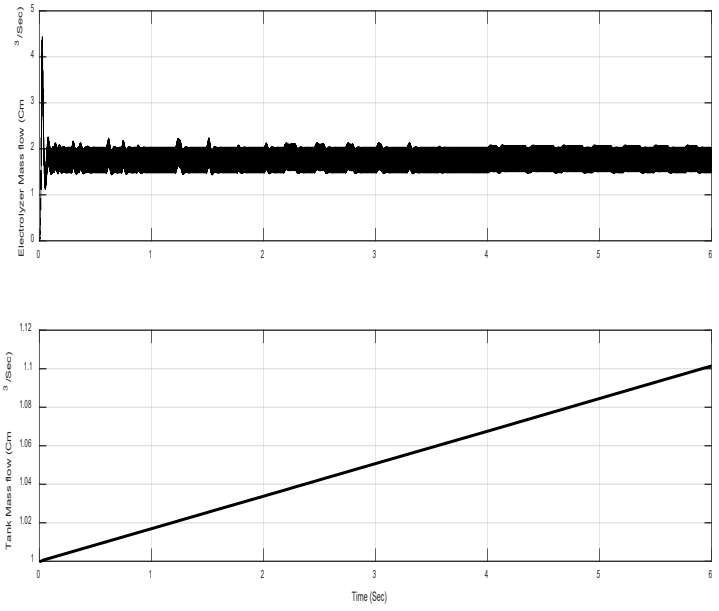


Figure 25. Electrolyzer mass flow and tank mass flow.

References

[1] Sarita Samal & Prakash Kumar Hota "Design and analysis of solar PV-fuel cell and wind energy based microgrid system for power quality improvement" Cogent Engineering, Vol. 4, No. 1, 2017. DOI :10.1080/23311916.2017.1402453

[2] S. Kumar Singh, C. Banerjee, "Modelling and Simulation of Hybrid Solar Cell and Fuel Cell for Suburban Electrification", IJAREEIE, Vol. 7, Issue 5, May 2018. DOI :10.15662/IJAREEIE.2018.0705040

[3] Ates, S. Z. Shekardasht, E. Canli "Solar Energy Supported Hydrogen Production: A Theoretical Case Study", Selcuk University, Selcuklu, Konya, TURKEY, 2017. DOI: 10.15317/Scitech.2017.110

[4] Mehran Jamshidi and Alireza Askarzadeh, "Techno-economic analysis and size optimization of an off-grid hybrid photovoltaic, fuel cell and diesel generator system" Sustainable Cities and Society, Vol. 44, No. 1, 2019, pp 310-320, DOI : 10.1016/j.scs.2018.10.021

[5] Juan M.Lujano-Rojas, Cláudio Monteiro, Rodolfo Dufo-López, and José L.Bernal-Agustín, "Optimum load management strategy for wind/diesel/battery hybrid power systems" Renewable Energy, Vol. 44, Aug. 2012, pp 288-295, DOI : 10.1016/j.renene.2012.01.097.

[6] Sumon Rashid, S. Rana, S.K.A. Shezan, Sayuti A.B. Karim, and Shamim Anower "Optimized design of a hybrid PV-wind-diesel energy system for sustainable development at coastal areas in Bangladesh" Environmental Progress & Sustainable Energy, Vol. 36, No. 1, 2017, pp 297-304, DOI : 10.1002/ep.12496.

[7] P. K. Goel, B. Singh, S. S. Murthy, N. Kishore, "Isolated wind-hydro hybrid system using cage generator and battery storage", IEEE Transactions on Industrial Electronics, Vol. 58, pp. 1141-1153, 2011. DOI: 10.1109/TIE.2009.2037646

[8] Iman Edwar, and Maged. N. F. Nashed "Solar and Wind Energy for Hybrid Electric Vehicle Controlled with Bi-directional DC/DC Converter System" 3rd International Conference on Advaced Control Circuits and Systems, (ACCS'013), Luxor, Egypt, Nov. 30 - Dec. 3, 2013.

[9] T. Bogaraj, J. Kanakaraj, and J. Chelladurai, "Modeling and simulation of stand-alone hybrid power system with fuzzy MPPT for remote load application," Archivals of Electrical Engineering, Vol. 64, No. 3, pp 487-504, DOI : 10.2478/aee-2015-0037, 2015.

[10] M. Idrees et al., "Fuzzy Logic Based Calculation and Analysis of Health Index for Power Transformer Installed in Grid Stations," in 2019 International Symposium on Recent Advances in Electrical Engineering (RAEE), 2019, vol. 4, pp. 1–6.

[11] Ehab M. Zakaria, Shawky H. Arafa, Maged M. N. F. Nashed, and Salah G. Ramadan "Dynamic Analysis of Photovoltaic-Fuel Cell-Electrolyzer System at Stand Alone Operation" International Journal of Engineering and Innovative Technology (IJEIT) Volume 8, Issue 11, May 2019, pp: 7-12

[12] S. Z. Hassan, H. Li, T. Kamal, F. Mehmood, "Fuzzy embedded MPPT modeling and control of PV system in a hybrid power system" 12th International IEEE Conference on Emerging Technologies (ICET) 18-19-October 2016, DOI: 10.1109/ICET.2016.7813236

[13] T. Kamal, S. Z. Hassan , M. Espinosa-Trujillo, H. Li, M. Flota, "An Optimal Power Sharing and Power Control Strategy of Photovoltaic/ Fuel Cell/Ultra-capacitor Hybrid Power System" Journal of Renewable and Sustainable Energy Vol. 8, No. 3, 2016. DOI: 10.1063/1.4948926

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 431-435 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Optimization of the Stabilizer Bar by Using Total Scores Method

Nguyen Tuan Anh[*,1], Hoang Thang Binh[2], Tran The Tran[2]

[1]*Mechatronics-Automobile Department, Hanoi University of Business and Technology, Viet Nam*

[2]*School of Transportation Engineering, Hanoi University of Science and Technology, Viet Nam*

ARTICLE INFO

Article history:
Received: 01 January, 2020
Accepted: 05 February, 2020
Online: 25 February, 2020

Keywords:
Stabilizer bar
Optimization
Thickness
Mass
Displacement

ABSTRACT

The stabilizer bar restricts the roll angle of the vehicle when the vehicle steers based on the load balance on both sides of the wheel. Usually, the bar is in the form of the symmetrical U-shaped with the diameter is circular or annular. In order to be able to calculate the diameter of the stabilizer bar, it is necessary to determine the value of force F acting on the arm of the stabilizer bar. Previous studies often estimate the force value F is a defined constant, thus causing inaccuracy in the calculation process. This study focuses on accurately determining the maximum force value F to optimize the size of the bar. The size of the bar is optimized based on the values of the mass and the displacement. The results of the study indicate that the bar with the annular section will be more optimal in terms of mass and displacement than the circular section. In most cases, if the thickness of the bar decreases, the mass and the displacement of the bar also tend to decrease.

1. Introduction

When the vehicle steers, under the influence of centrifugal force will cause the vehicle body to be tilted and change the load on both sides of the wheel [1, 2]. The stabilizer bar is used to restrict the roll angle of the vehicle when steering, redistribute the load on both sides of the wheel, thereby supporting the vehicle to move more stably and safely [3, 4].

The stabilizer bar in Figure 1 has a structure consisting of two parts: the back and the arm. Normally, the stabilizer bar is in the form of symmetrical U-shaped with constant diameter along the length, some bars can be hollowed out to reduce mass [5–7]. The size and geometric parameters of the bar depend on the position of the bar layout on the vehicle.

When the load on both sides of the wheel changes, the arm will move in opposite directions, causing opposite torque on the back of the bar. These moment values are similar, so they cancel each other out and make the vehicle more stable.

Previous studies often focused on calculating the stresses and displacements of bars. In order to be simplified during the calculation process, the force value F acting on the arm of the stabilizer bar is generally estimated [8–12]. However, this method will cause major disadvantages as follows:

If the value of the force F is too small, the diameter of the bar will be small. Therefore, special moving conditions may affect the durability and reliability of the bar. Besides, when the vehicle is tilted too large, it will cause the arm's displacement to exceed the limited space.

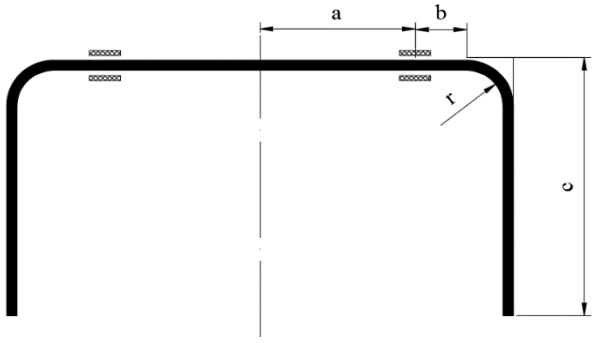


Figure 1: The geometry of the stabilizer bar.

If the value of force F is too large, the diameter of the bar will be too large. Therefore, it will affect the layout of the bar on the vehicle. Besides, if the size of the bar is too large, the unsprung mass will increase, causing a loss of comfort for the passenger.

This study focuses on optimizing the size of the stabilizer bar based on the exact determination of the force value acting on

* Nguyen Tuan Anh, Email: taae177@gmail.com

the arm. Additionally, the study also establishes the dependence between the roll acceleration and the vehicle's mass and velocity.

2. Determine the diameter of the stabilizer bar

2.1. Requirements for the stabilizer bar

With the function of limiting the roll angle of the vehicle body when steering, the stabilizer bar on the vehicle needs to meet some requirements when designing such as:

- Generate anti-roll moments large enough to limit the roll angle.
- Ensuring torsional and bending conditions during operation (especially for large size vehicles).
- The geometry and displacement must be reasonable.
- The mass is small and limited to affect the comfort of the vehicle.

2.2. Determine the force value acting on the stabilizer bar

The force value acting on the arm of the stabilizer bar can be determined as follows: [13]

$$\Delta F = \left| F_i - \frac{mg}{2} \right| \quad (1)$$

When the vehicle body is tilted, the trajectory of the center is an arc with a radius is the distance from the center of gravity to the roll center [14]. The maximum force value acting on the arm of the stabilizer bar will be:

$$F_{max} = \frac{mh(gtan\varphi_{max} + a_{ymax})}{2t} \quad (2)$$

where

F_i: The force of the suspension system.

φ_{max}: The limited roll angle of the vehicle body.

a_{ymax}: The limited lateral acceleration of the vehicle (respectively to φ_{max}).

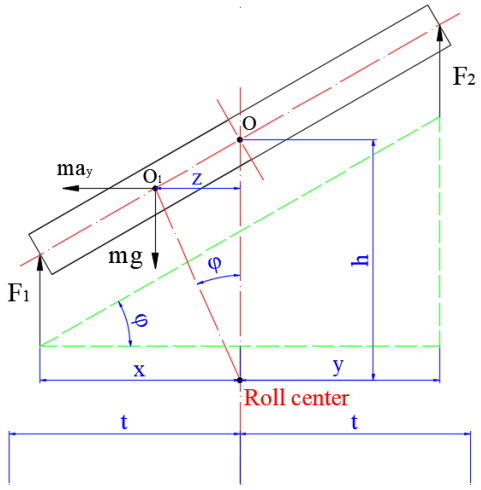


Figure 2: The force distribution diagram acting on the vehicle body.

Assume that: u is the roll acceleration:

$$u = gtan\varphi_{max} + a_{ymax} \quad (3)$$

The formula (2) can be rewritten as:

$$F_{max} = \frac{mhu}{2t} \quad (4)$$

From (4) can see that:

- The load acting on the stabilizer bar is proportional to the mass and dimension of the vehicle. However, the values of mass and dimension are unchanged when the vehicle moves on the road.
- Load acting on the stabilizer bar depends on roll angle and lateral acceleration of the vehicle. These two values vary according to the velocity, steering angle, road surface conditions and other effects of external forces.

Therefore, in order to determine the maximum load value acting on the arm part of the stabilizer bar, the limit values φ_{max} and a_{ymax} need to be determined. However, in different conditions, the limit of roll angle and lateral acceleration will be different, it is difficult to accurately determine these values. Normally, there will be 2 cases of interest:

Case 1: Vehicle steers unstable (steering angle is large, steering acceleration is high) at a high speed, the vehicle will rollover very fast while φ_{max} and a_{ymax} are still not big enough.

Case 2: Vehicle steers stable (steering angle is small) at high speed, the value of φ_{max} and a_{ymax} will be larger.

Therefore, the maximum force value acting on the arm of the stabilizer bar will be determined according to the second case.

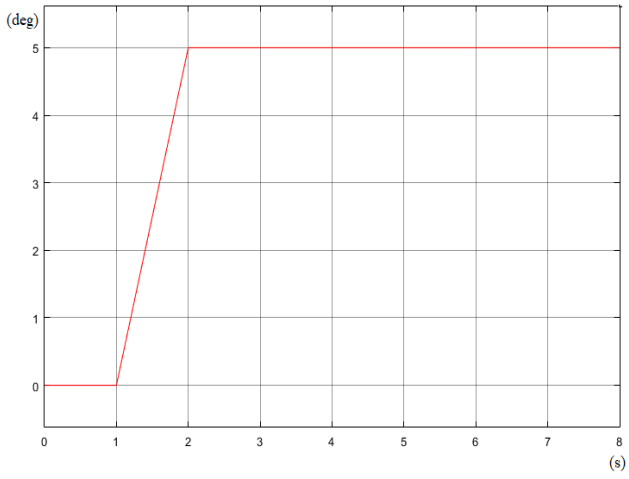


Figure 3: Steering angle.

Using the established dynamic vehicle model [15], reference vehicle parameters such as Table 1 and steering angle as shown in Figure 3. The graph shows the dependence of the roll acceleration on the mass and velocity of the vehicle is given as shown in Figure 4.

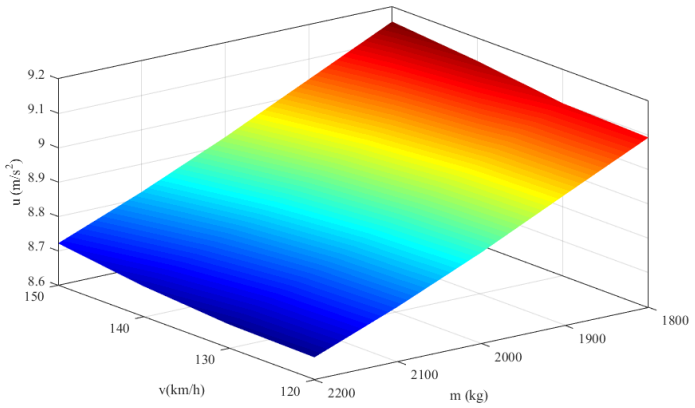


Figure 4: The dependence of the roll acceleration on mass and velocity.

From the graph in Figure 4, it can be seen that:

- When the sprung mass increases, the value of the roll acceleration tends to decrease.
- When the speed increases, the value of the roll acceleration increase.

Therefore, the force value F acting on the arm of the stabilizer bar is taken in case the vehicle moves at high velocity.

2.3. Calculation of the stabilizer bar

The stabilizer bar must ensure torsional and bending conditions during work. The diameter of the stabilizer bar is determined by following the below formula [16]:

$$D \geq \sqrt[3]{\frac{16F\sqrt{4(b+r)^2+3c^2}}{\pi[\sigma]}+d^3} \quad (5)$$

Where: D is the outer diameter and d is the internal diameter of the stabilizer bar.

After determining the diameter of the stabilizer bar. The mass of the stabilizer bar can be calculated based on the formula (6):

$$m_{SB} = \frac{1}{2}\rho\pi\left(D^2-d^2\right)\left(a+b+c+\frac{\pi r}{2}\right) \quad (6)$$

According to [17], the displacement of the stabilizer bar can be determined as follows:

$$\delta = \frac{F(b+r)^3}{3EJ_x} + c\tan\beta \quad (7)$$

with β is the torsional angle of the stabilizer bar:

$$\beta = \int_0^{a+b} \frac{Fc}{GJ_p}dz \quad (8)$$

2.4. Optimization of the stabilizer bar

Use reference parameters as shown in Table 1.

The simulation performed at the velocity value v = 150 km/h, the value of the roll acceleration is determined from the Figure 4. The diameter, mass, and displacement of the stabilizer bar are given as shown in Table 2.

From Table 2, it was found that:

- For solid bars, the outside diameter is small but the mass and displacement of the bar are very large.
- For hollow bars, the mass and displacement will vary depending on the thickness of the bar.
    - If the stabilizer bar is optimized in terms of total scores, the optimal value can be used as d_1 = 24 (mm); D_1 = 30 (mm) or d_2 = 25 (mm); D_2 = 31 (mm).

Table 1: Reference parameters.

Description	Symbol	Value	Unit
Sprung mass	m	2000	kg
Distance from center of gravity to front axle	l_1	1300	mm
Distance from center of gravity to rear axle	l_2	1700	mm
Distance from bearing to center	a	300	mm
Distance from bearing to bending angle	b	130	mm
Length of the arm	c	250	mm
The radius of curvature	r	70	mm
Limited stress	σ	1200	MPa
Density	ρ	7800	kg/m^3

Table 2: The size of the stabilizer bar.

Inner diameter (mm)	Outer diameter (mm)	Mass (kg)	Score 1	Displacement (mm)	Score 2	Total score
0	23	5.12	1	61.64	1	2
10	24	4.61	2	53.39	4	6
11	24	4.40	4	54.19	3	7
12	24	4.18	5	55.28	2	7
13	25	4.41	3	47.33	7	10
14	25	4.15	6	48.68	6	12
15	25	3.87	9	50.46	5	14
16	26	4.07	7	43.72	9	16
17	26	3.75	10	45.86	8	18
18	27	3.92	8	40.08	12	20
19	27	3.56	12	42.65	10	22
20	28	3.72	11	37.56	14	25
21	28	3.32	14	40.69	11	25
22	29	3.46	13	36.09	15	28
23	29	3.02	17	39.98	13	30
24	30	3.14	16	35.69	16	32
25	31	3.25	15	31.99	17	32

3. Conclusion

The stabilizer bar restricts the roll angle when the vehicle steers, and evenly distributes the load on both sides of the wheel, assisting a smooth wheel on the road.

The size of the stabilizer bar greatly affects the mass and displacement of the bar. If the bar is thinner and smaller, the mass and displacement of the stabilizer bar to decrease. Therefore, the comfort of the vehicle will be increased.

The stabilizer bar optimization process to find the most suitable dimension value. Therefore, it can reduce the mass and save materials in the production process.

The results of the study indicated that: With the vehicle specification used, the optimum size of the stabilizer bar was chosen as $d_1 = 24$ (mm); $D_1 = 30$ (mm) or $d_2 = 25$ (mm); $D_2 = 31$ (mm). In the case of the mass of the stabilizer bar is optimized, the size d_1; D_1 should be used. However, if the displacement of the stabilizer bar is optimized, the size d_2; D_2 should be used.

This research has only provided the optimal theoretical calculation method. Experimentation is needed to get the best effect.

References

[1]. N. K. Trai, N. T. Hoan, H. H. Hai, P. H. Huong, N. V. Chuong, T. M. Hoang, Structure Vehicle, Bach Khoa Hanoi Publishing, 2010.

[2]. R. Rajamani, Vehicle Dynamics and Control, Springer Publishing, 2012. DOI: 10.1007/978-1-4614-1433-9

[3]. A. N. Khartode, M. U. Gaikwad, "Design, and Analysis of Anti-roll Bars for Automotive Application", International Journal on Recent and Innovation Trends in Computing and Communication, 4(6), 340–345, 2016.

[4]. R. Stone, J. K. Ball, Automotive Engineering Fundamentals, SAE International Publishing, 2004.

[5]. H. Bankar, R. Kharade, P. Baskar, "Finite Element Analysis of Anti-roll Bar to Optimize the Stiffness of the Anti-roll Bar and the Body Roll", International Journal of Model Engineering Research, 3(10), 11–23, 2014.

[6]. S. Ramakrishna, B. R. H. Reddy, B. Akhil, B. P. Kumar, "A Review on Anti-roll Bar Used in Locomotives and Vehicles", International Journal of Current Engineering and Technology, 7(3), 838–841, 2017.

[7]. A. P. P. Mariappa, Design and Optimization of Anti-roll Bar, Master Thesis, Kaunas University of Technology, 2018.

[8]. S. B. Tuljapure, L. S. Kanna, "Analysis on Stability Bar", Advanced Materials Manufacturing & Characterization, 3(1), 349–354, 2013.b

[9]. P. Bharane, K. Tanpure, A. Patil, G. Kerkal, "Design, Analysis, and Optimization of Anti-roll Bar", International Journal of Engineering Research and Applications, 4(9), 137–140, 2014.

[10]. P. G. Jamode, J. S. Gawande, T. S. Hingve, "Structural Analysis and Validation of Anti-roll Bar", International Journal of Interdisciplinary Innovative Research and Development, 1(4), 72–74, 2017.

[11]. K. K. Sharma, A. Rashid, S. Mandale, "Analysis of Anti-roll Bar to Optimize the Stiffness", International Journal of Modern Trends in Engineering and Research, 2(7), 1874–1879, 2015.
[12]. A. M. Wittek, D. Gaska, B. Lazarz, T. Matyja, "Automotive Stabilizer Bar–Stabilizer Bar Strength Calculations Using FEM, Ovalization of Radial Areas of Tubular Stabilizer Bar", Mechanika Journal, 20(6), 535–542, 2014. DOI: 10.5755/j01.mech.20.6.7706
[13]. N. T. Anh, "Study on the Load Distribution of the Vehicle When Steering at Different Velocities", International Journal of Innovative Science and Research Technology, 4(5), 433–435, 2019.
[14]. N. T. Anh, H. T. Binh, "Study on the Influence of Vehicle's Dimension to Rollover", International Journal of Science and Engineering Investigations, 8(5), 62–64, 2019.
[15]. N. T. Anh, H. T. Binh, "Research on Determining the Limited Roll Angle of Vehicle", Advances in Engineering Research and Application, Lecture Notes in Network and Systems, vol. 104, 613–619, 2020. DOI: 10.1007/978-3-030-37497-6_70
[16]. T. T. Hung, Strength of Materials, Science and Technology Publishing, 2006.
[17]. N. T. Anh, T. T. Tran, H. T. Binh, P. H. Nam, L. T. Dung, "The Study on the Method of Calculating and Designing the Stabilizer Bar on the Vehicle Using Solidworks Software", Vietnam Mechanical Engineering Journal, 7(12), 92–99, 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 436-445 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Seismic Response Modification Factor for Special Concrete Structures Based on Pushover &Non-linear Time History Analyses

Rasha Attia Ahmed Attia*, Reham Mohamed Galal Ebrahim El-Tahawy, Mohamed Nour El-Din Fayed

Structural Department, AinShams University, 11311, Cairo, Egypt

ARTICLE INFO

Article history:
Received: 01 January, 2020
Accepted: 12 February, 2020
Online: 25 February, 2020

Keywords:
Ductility Factor
Pushover analysis
Reduction Modification Factor
Time history analysis

ABSTRACT

Reinforced concrete (RC) structures are able to dissipate a large amount of energy during earthquake (EQ). RC structures designed to resist EQs must have enough strength and stiffness to prevent any possible collapse as well as to control deflection. Seismic codes include the response reduction/modification factor(R) in calculation of equivalent lateral forces used for EQ resistant buildings design to reduce elastic design spectral acceleration of the RC structure to account for ductility and over strength. Furthermore, the aim of this research is to investigate the variation of seismic response of special RC structures with different configurations, such as elevated metro stations. Consequently, the response reduction factor can be evaluated. Then the calculated response reduction/modification factors (R) for reinforced concrete (RC) structures will be compared to those specified in ECP and the ASCE code. For this purpose, a 3D Finite Element Method was used for modelling RC structure using ABAQUS V.6.14 program.

1. Introduction

Experience in the past EQs has demonstrated that many typical methods of construction lack basic resistant to EQ forces, so that the concept of EQ design has been developed [1]. The basic approach of EQ design should depend on lateral strength, ductility and deformability capacity of reinforced concrete (RC) structure with specific level of damage while preventing collapse. Therefore, one of the primary tasks of the structural engineer designing an earthquake resistant building is to ensure that the RC building possess enough ductility and strength to resist the expected levels of EQs during its lifetime.

In design codes, Non-linearity of structures in linear analysis is indirectly considered by reducing the seismic demand on RC structures or a (R) factor proportional to the expected ductility of structures. Pushover & non-linear time history analyses could also be performed for better estimate for EQ forces [2].

2. literature Review

A summary of previously available work was conducted on different fields like response reduction/modification factor(R), displacement ductility capacity, ductility, pushover and time history for this study is presented as follows.

- In 2016, The author [3] modeled RC SMRF with medium rise that have irregularity in elevation& plan to evaluate the (R) factor for irregular RC structures using the non-linear static analysis. The results explained that increasing in horizontal & vertical irregularities percentage leads to a smaller value for the (R) factor.
- In 2015, The author [4] modeled two reinforced concrete(RC) frames types, (SMRF) & (OMRF) with a constant number of bays and number of stories (regular RC frame and soft storied frame), mass irregular frame and geometric irregular frames. The results explained that both (SMRF) and (OMRF) failed to reach the target values of (R) factor specified in Indian standards code.
- In 2014, The author [5] evaluated (R) factors of frames from their pushover curves to check its adequacy if compared to the R-value recommended in Indian standards (IS). RC frames with number of stories varying from two to twelve with four bays were analyzed. The aspect ratios of frames evaluated were not the same. The results explained that, over strength factor decreases with increasing in number of stories, (R) factor decrease with increase in a number of stories and Indian standards (IS) code giving conservative value of (R) factor for regular RC frames structures.

*Corresponding Author Name: Rasha Attia Ahmed, rasha.attia@hotmail.com

- In 2009, The author [6] modeled two RC structures with varying configurations to estimate (R) factor for the RC frames in Pakistan. This research included a different analysis including incremental dynamic analysis and inelastic static pushover. Seven natural records of EQ were selected. This study explained that the analysis of the proposed (R) factor in codes gives a false representation of the structure response during EQs. Also the results explained that the (R) factor mentioned in BCP2007 (UBC 97) or the NESPAK2006 become un-conservative by up to 11% and 22%, respectively.
- In 2008, The author [7] proposed the use of pushover analysis as a method to specify the performance level of a RC structure designed according to the Algerian code. Since behavior of RC structures could be highly inelastic if subjected to lateral loads, plastic yielding effects would conquer the total in-elastic performance of RC constructions.

3. Response Reduction Factor

The (R) factor is a measure of the RC structure ductility and over strength in inelastic phase. It also can be expressed as a function of different parameters of RC structural system as damping, ductility, redundancy, and strength as per ATC-19 [8 & 9].

$$R = R_S R_\mu R_R R_\xi \quad (1)$$

,where (R_s): over strength factor, (R_μ): ductility factor, (R_ξ): damping factor.

- The over-strength factor (R_s) is defined in ASCE 7(2008) by the ratio between (V_u) & (V_d), where the design lateral load equal 60% of (V_u) to satisfy the system requirement to remain elastic as suggested by Uang (1991) [10]. Design displacement (Δ_d) is defined as the displacement corresponding to the design load.

- (R_μ): Ductility factor is a function of μ, for short, intermediate and long period structures is presented below in equation (2) as suggested by Newmark and Hall (1982) [11]:

Short period T < 0.2 seconds R_μ=1

Intermediate period 0.2<T<0.5 seconds

$$R\mu = \sqrt{2\mu - 1} \quad (2)$$

Long period T > 0.5 seconds R_μ=μ

- (R_ξ): Damping factor balances the effect of supplementary viscous damping and is mainly applicable in case of structures with additional energy dissipating devices. In the absence of such devices, the damping factor is generally assumed as 1.0.
- (R_R): Redundancy factor depends on the number of vertical elements participating in seismic resistance and can be assumed as unity as per ASCE7 guidelines.

Figure 1 illustrate relationship between (R) factor, over-strength factor (R_s) and ductility factor R_μ; taking R_R & R_ξ =1

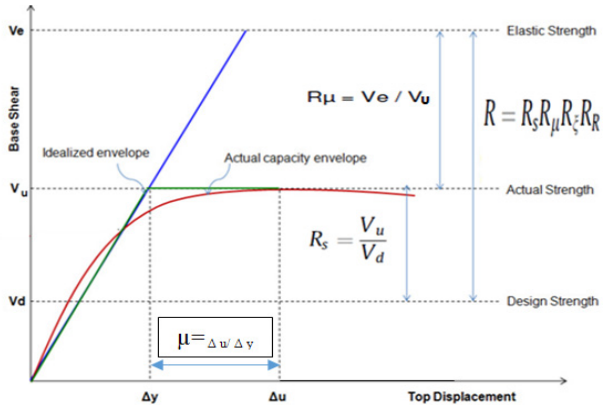


Figure 1: Relationship between (R) factor, (R_s) & (R_μ)

3.1. Response Reduction Factor (R) in ECP and Other International Codes

The (R) factor is assumed to represent ratio between forces that would develop under lateral loads if the framing system was to behave entirely elastically to the prescribed design forces at the strength level. (R) factors are used in current building codes to estimate RC structures strength designed using linear and nonlinear methods. Use of (R) factor has been clarified in the seismic design codes in many parts of the world, which are affected by earthquakes. A brief overview of (R) factor in the design seismic codes of ECP [12] and ASCE [13] for special RC structure is illustrated in the following tables.

i. ASCE (American Society of Civil Engineers)

(R) factor values for RC structures recommended in ASCE-7-10, the (R) defined for cantilever beam supported on RC column systems detailed to conform to the requirements for moment resisting frames, based on special or intermediate or ordinary MRF systems these values are presented in Table 1.

Table 1: R factor in ASCE code

R factor in ASCE7-10code		
Structural Systems	Type	R
(RC)Moment resisting frame	Special	8
	Ordinary	3
	Intermediate	5

Table 2: R factor in ECP code

R factor in ECP code		
Structural Systems	Type	R
(RC)Moment resisting frame	Sufficient ductility	7
	Limited ductility	5

ii. ECP (Egyptian Code of Practice)

ECP code doesn't specify a unique value for the response reduction factor for special structures like elevated metro stations where the structural system consists of a single column with

double cantilevers supporting a multi-story reinforced concrete building. But a table showing R factor for sufficient and non-sufficient can be used in comparison as shown in Table 2.

iii. UBC (Uniform Building Code)

Values of (R) factor for RC structures recommended in UBC code (1997) [14] are presented in Table3.

Table 3: R factor in UBC code

R factor in ASCE-7-10code		
Structural Systems	Type	R
(RC)Moment resisting frame	Special	8.5
	Ordinary	3.5

4. Non-Linear Numerical Model For RC Structures

4.1. Element Model

ABAQUS V.6.14 program [15] is a finite element package for structural analysis, able to simulate behavior of different types of structures and materials by elements which can be adjusted to fit the geometry of the model. The program offers a wide range of capabilities to solve complex problems in mechanics.

4.2. Material Model

ABAQUS (Hibbitt, Karlsson and Sorensen, 2001) supplies the Concrete Damaged Plasticity (CDP) model for dynamic loading. Tensile & compressive response of concrete can be characterized by CDP in Figure 2.

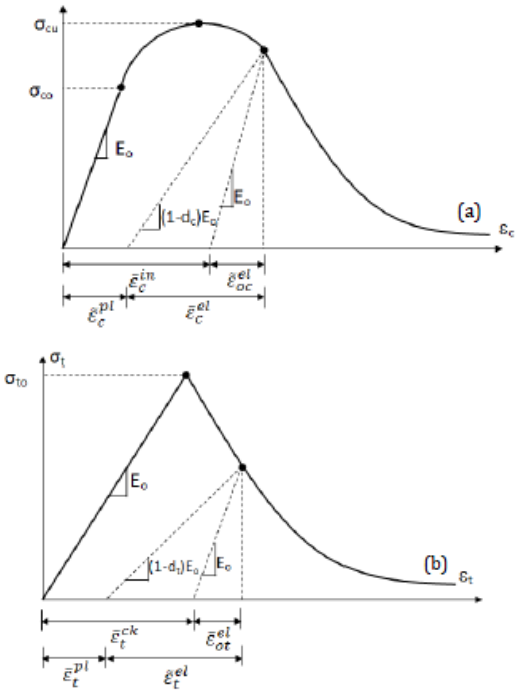


Figure 2. Behavior of concrete under axial compressive (a) and tension (b) strength (ABAQUS User Manual, 2008 [16])

As shown in Figure 2, the unloaded response of concrete specimen seems to be weakened because the elastic stiffness of the material appears to be damaged or degraded. The degradation of the elastic stiffness on the strain softening branch of the stress-strain curve is characterized by two damage variables, d_t and d_c, which can take values from zero to one. Zero represents the undamaged material where one represents total loss of strength (ABAQUS User Manual, 2008). E_0 is the initial (undamaged) elastic stiffness of the material and $\varepsilon c^{\sim pl}$, $\varepsilon_t^{\sim pl}$, $\varepsilon c^{\sim in}$, $\varepsilon_t^{\sim in}$ are compressive plastic strain, tensile plastic strain, compressive inelastic strain and tensile inelastic strain respectively. The stress-strain relations under uni-axial tension and compression are taken into account in Eq. (3) and Eq. (4).

$$\sigma_t = (1 - d_t) E_O (\varepsilon_t - \varepsilon_t^{\sim pl}) \quad (3)$$

$$\sigma_c = (1 - d_c) E_O (\varepsilon_c - \varepsilon_c^{\sim pl}) \quad (4)$$

An exponential function has been used to calculate d_o compression & tension damage, as follows:

$$d_c = 1 - \exp^{(-ac\ \varepsilon c^{in})} \quad (5)$$

$$d_t = 1 - \exp^{(-at\ \varepsilon t^{ck})} \quad (6)$$

a_t and a_c are model parameters for uni-axial tension & compression, and can be calibrated from the following conditions:

$$\varepsilon_c^{pl} = 0 \rightarrow d_c = 0 \ \& \ \varepsilon_c^{pl} = \varepsilon_c^{pl}{}_{max} \rightarrow d_c = 1 \quad (7)$$

$$\varepsilon_t^{pl} = 0 \rightarrow d_t = 0 \ \& \ \varepsilon_t^{pl} = \varepsilon_t^{pl}{}_{max} \rightarrow d_t = 1 \quad (8)$$

4.3. Model Characterization

The goal is to establish a simple 3D nonlinear model for RC model. The model employed element C3D20, which is a nonlinear 20-node solid element for concrete and steel rebar's.

The elements were connected together with appropriate constraints to represent the interaction between various components of the model assembly.

Element size has significant effect on accuracy of finite element results. Two types of mesh elements were used; linear and quadratic, Figure 6 shows that using quadratic meshing made the error smaller due to Small size of FE leads to the smaller error with less computational time.

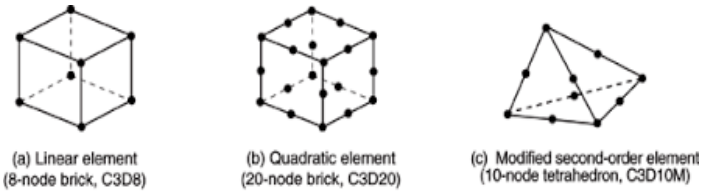


Figure 3: ABAQUS shell element

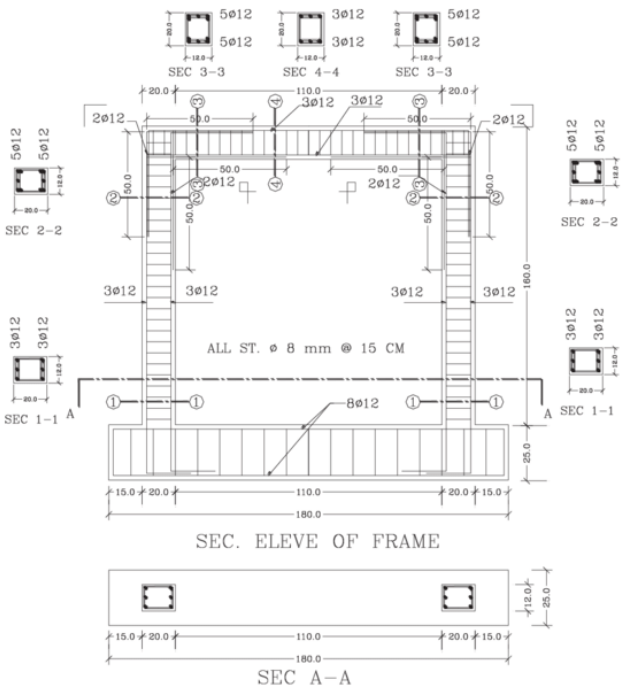
5. Verification Examples

Two experimental output data for reinforced concrete (RC) frames using pushover and one using time history in order to check reliability and validity of a nonlinear finite element model.

For the numerical Pushover investigation, two reinforced concrete frame specimens experimentally tested by Lila M.Abdel-Hafez et.al. (2015) [17] and Ali Mansouri et al. (2013) [18] under lateral loading were modeled. Figure 4 presents the cross-section details of the tested RC frames and their dimensions. Figure 5 shows the ABAQUS analysis models developed for the tested models respectively. While Figure 6 shows the plastic hinges formation and crushing in concrete for both models.

5.1. Experimental models

1. Lila M.Abdel-Hafez et.al. (2015)



2. Ali Mansouri et al. (2013)

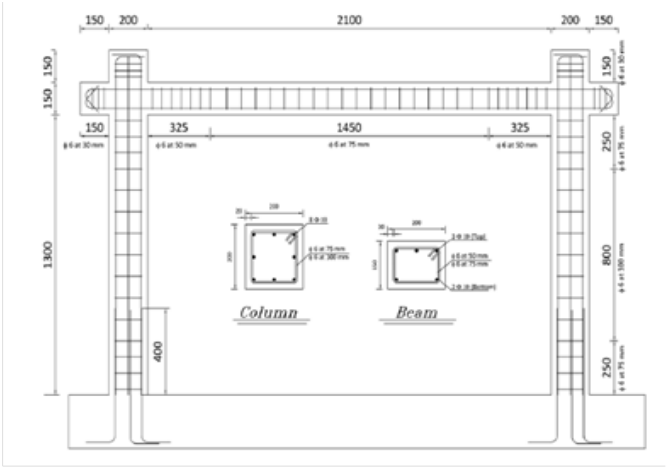


Figure 4: Cross-Sections of the tested RC frames and their dimensions

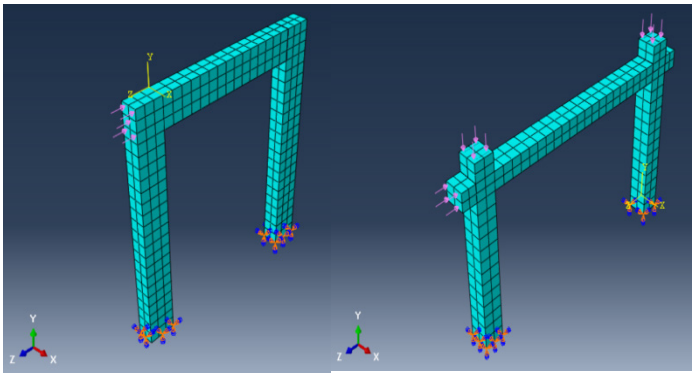


Figure 5: ABAQUS analysis models developed for the tested RC frames

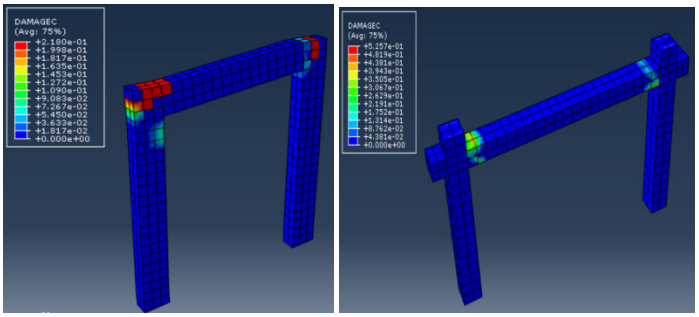


Figure 6: Plastic hinges formation & crushing in concrete

3. Aslan Sadeghi Hokmabadi et, al.(2014)

For the numerical nonlinear time history analysis investigation, three records for each were compared by Aslan Sadeghi Hokmabadi (2014) [19]. Figure 7 illustrate the shaking table tests were performed by applying scaled earthquake acceleration records of 1994 Northridge, 1995 Kobe, and 1940 El Centro to the fixed-base structural models. Figure 8 shows the ABAQUS analysis model using time history analysis.

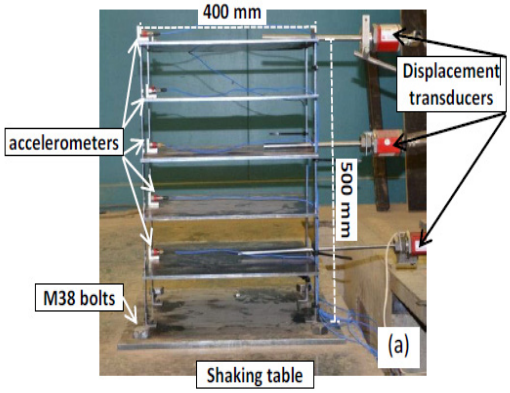


Figure 7: Five storey fixed-base model structure for shaking table tests

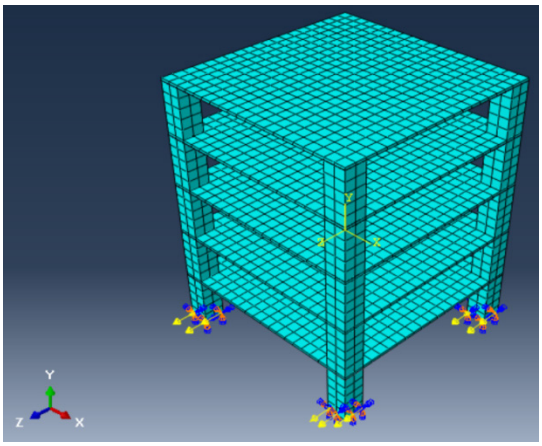


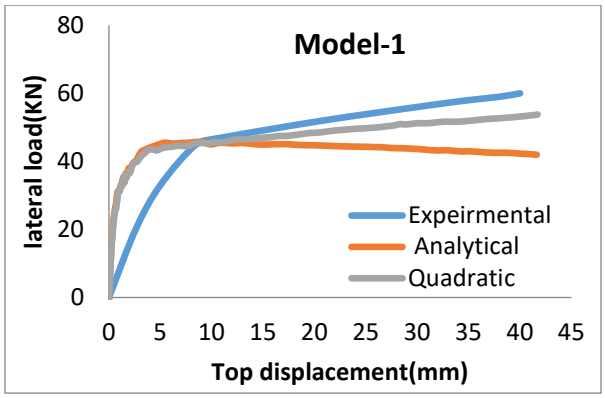
Figure 8: ABAQUS analysis model for Five storey fixed-base structure

5.2. Evaluation of the Numerical Results

Figure 9 and 10 show that there is a good agreement between the experimental results and the corresponding results from the pushover analysis for tested RC frames. A comparison between analytical & experimental results is presented regarding to pushover curve in Table 4, 5 and the error in results as shown is

taken as the ratio between the difference of analytical and experimental result to the experimental value.

$$\left(\frac{\text{Analytical mean value}-\text{Experimental mean value}}{\text{Experimental mean value}}\right) \text{x}100$$



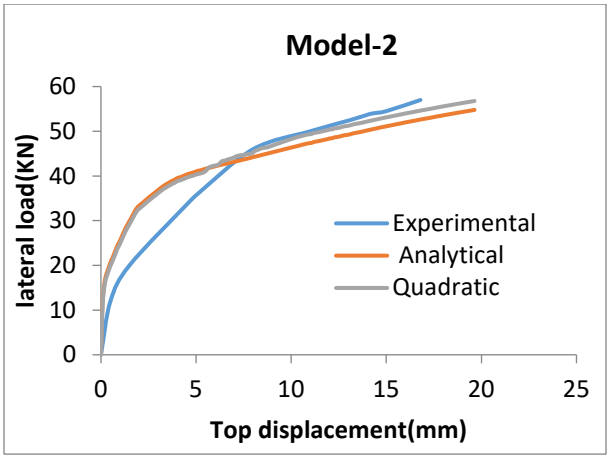


Figure 9: Modeling Results for the two reinforced concrete (RC) frames specimens

Table 4: Results for the two reinforced concrete (RC) frames specimens

Model	Experimental (shear) Qult(KN)	Analytical (shear) Qult(KN)	% Error
1	65	55	15.3
2	60	57	5.00

Model	Experimental (displacemen) Δult (mm)	Analytical (displacement) Δult (mm)	% Error
1	40.07	41.70	4.07
2	16.8	19.53	16.2

6. Applications for Special Structures, RC Elevated Metro Stations

In this study, RC elevated metro station supported on fixed base is utilized to estimate the response reduction/modification factor(R) using time history and static push-over analysis. Estimation of R value in this study depends significantly on how well the non-linear behavior of this RC system is presented in analysis. The selected studied RC metro station and it's details is shown in Figure 11 and 12. RC elevated metro station is designed according to ECP code for design 2017 [20] & ECP code for loading 2012 [12] and material properties are taken as following:

- Concrete compressive strength (cube) (fcu) =40 MPa
- Steel yield strength (fy) = 400 MPa for vertical reinforcement and stirrups.
- Steel young's modulus (Es) =200 GPa
- Specific weight for wall =18 kN\m3
- Dimension of concrete column (2000x3000) mm with height equal to 8m.
- Dimension of beam girder (2000x3750) mm.
- Height of platform level and mechanical level =2.80m
- Spacing between columns equal to 14m, total length of metro station equal to 42 m with height 21.9m.
- All RC sections were designed using ETABS software.

RC elevated metro station loads are considered according to ECP code of loads as follow:

- Considered loads for platform and under platform level, FC= 4 kN/m2 and LL= 6kN/m2.

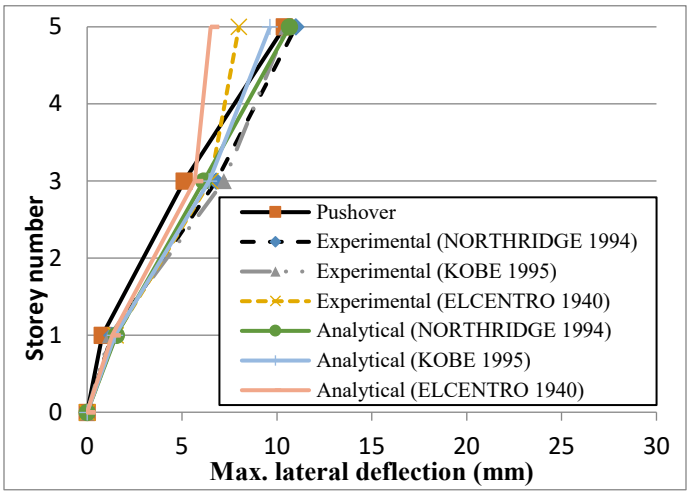


Figure 10: Comparison between pushover and time history records for Max. Lateral deflection (mm)

Table 5: Results for nonlinear time history analysis

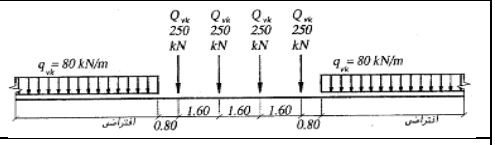
Story No.	(NORTHRIDGE 1994)		(KOBE 1995)		(ELCENTRO 1940)	
	Experimental	Analytical	Experimental	Analytical	Experimental	Analytical
0	0	0	0	0	0	0
1	1.3	1.55	1.2	1.473	1.5	1.32
3	6.8	6.14	7.2	6.43	6.5	5.67

% Error (NORTHRIDGE 1994)	
Story No.	Displacement
0	0
1	19.2
3	9.7
5	3.1

% Error (KOBE 1995)	
Story No.	Displacement
0	0
1	22.75
3	10.69
5	8.19

% Error (ELCENTRO 1940)	
Story No.	Displacement
0	0
1	12.00
3	12.77
5	18.63

Considered loads for box girder supporting the train according to ECP code in the following table:

Live Loads	
Sample model	
Pulling Force	300+24(Lo-7)KN
Braking Force	250+20((Lo-7)KN
Lateral Shock Effect	100 KN (in horizontal direction perpendicular to railway direction)

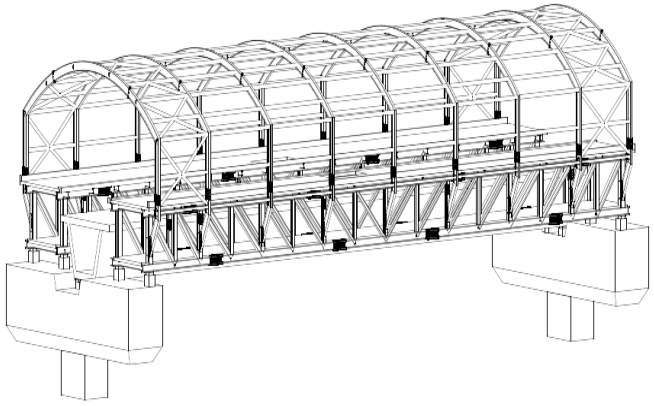
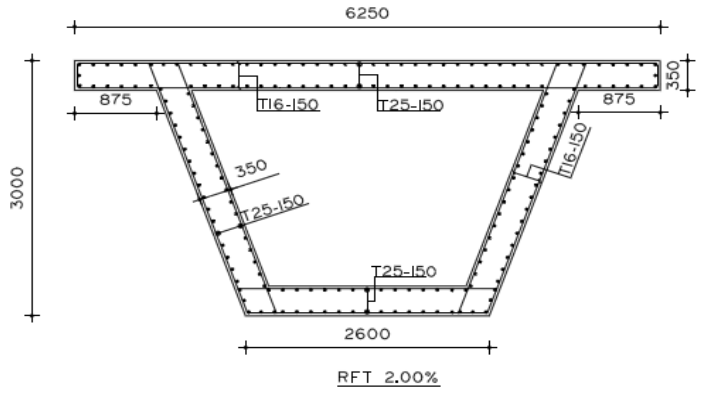
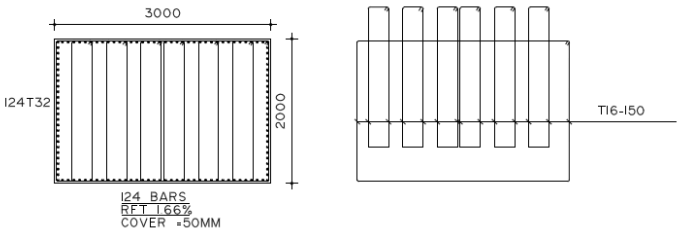


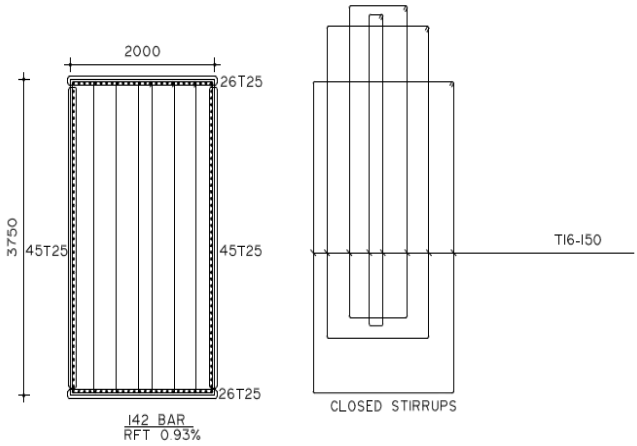
Figure 11: Elevated RC metro station system



TYPICAL DETAIL FOR BOX GIRDER
SUPPORTING ELEVATED METRO STATION



TYPICAL DETAIL FOR RC COLUMN
SUPPORTING ELEVATED METRO STATION



TYPICAL DETAIL FOR RC BEAM GIRDER
SUPPORTING ELEVATED METRO STATION

Figure 12: Elevated RC metro station

Figure (13) shows the ABAQUS modeling for concrete solid elements and steel wire elements, while Figure (14) shows the boundary conditions for the concrete footing had locked translational degrees of freedom in all directions.

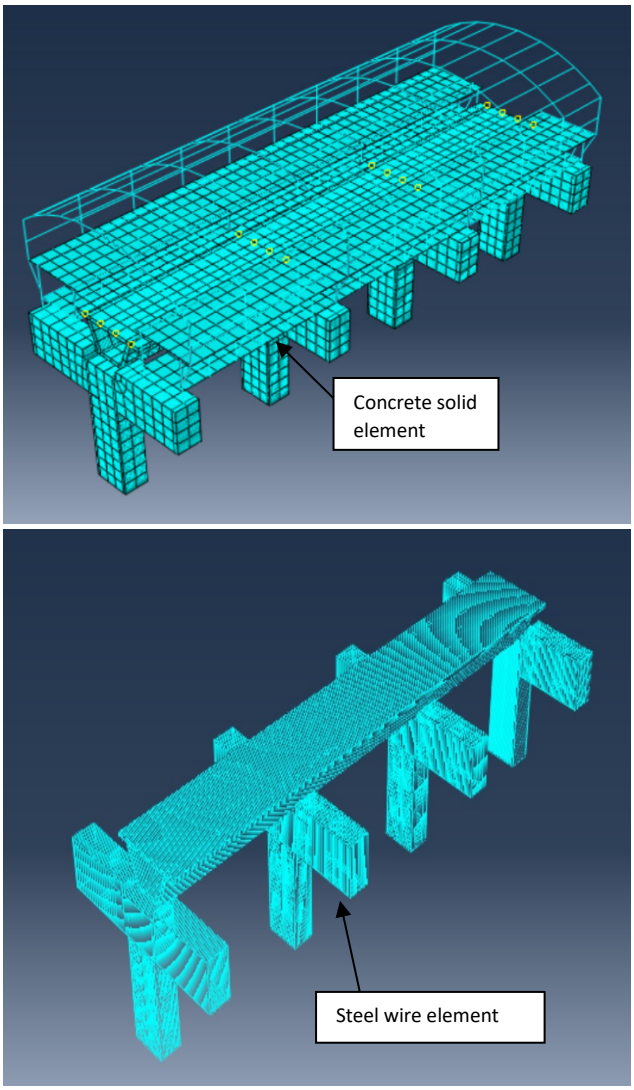


Figure 13: Element mesh

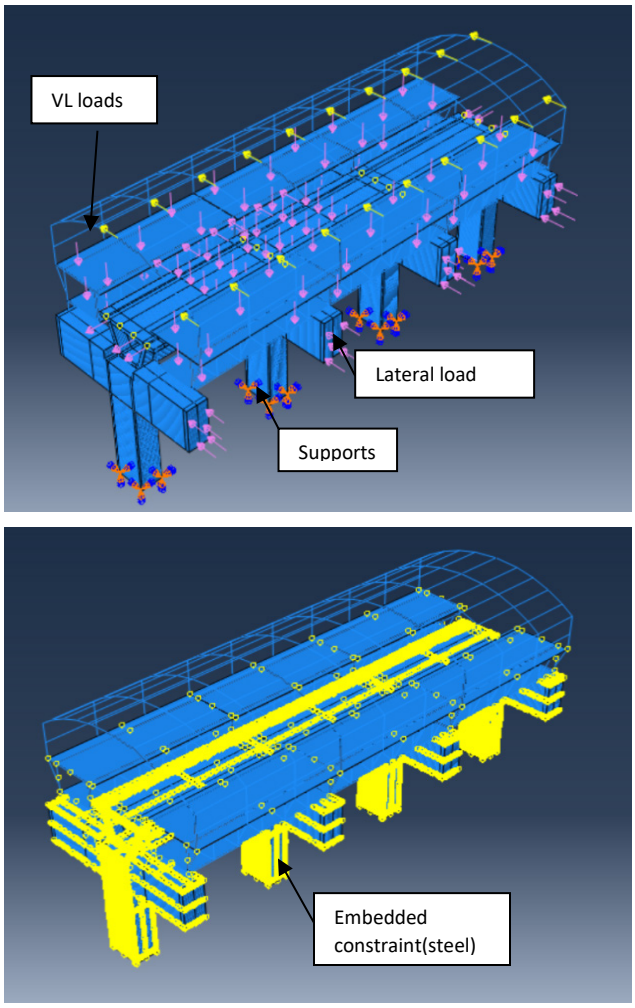


Figure14: Boundary conditions, loads and constraints

7. Cases of Study

Four configurations for RC elevated metro stations were analyzed as follows:

- 2 platform levels, considering under platform & platform levels with and without steel cladding, as shown in Figure 15.

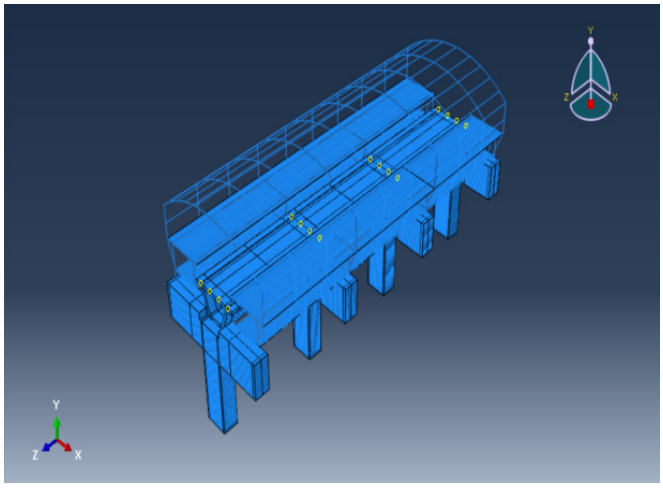


Figure 15.a: Two platform levels with steel cladding

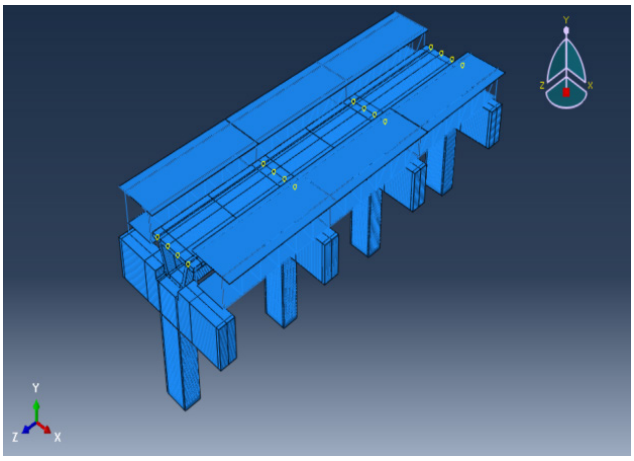


Figure 15.b: Two platform levels without steel cladding

- 3 platform levels, considering under platform, platform levels & mechanical level with and without steel cladding, as shown in Figure 16.

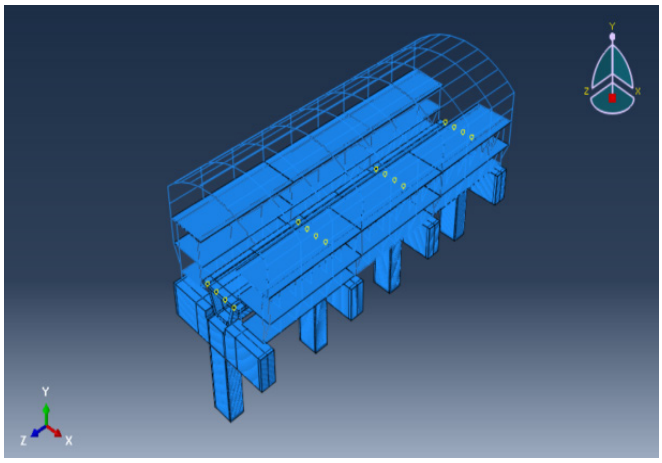


Figure16.a: Three platform levels with steel cladding

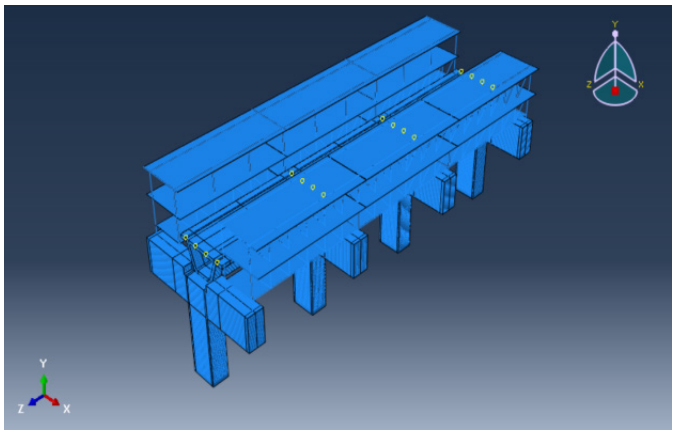


Figure16.b: Three platform levels without steel cladding

8. Failure Criteria

The analyses were continued until the stress in reinforcement reached 400MPa as shown in Figure 17 or the strain in concrete reached 0.003 (according to ECP code [12]) as shown in Figure 18. While Figure 19 shows plastic hinges formation at fixation and crushing in concrete.

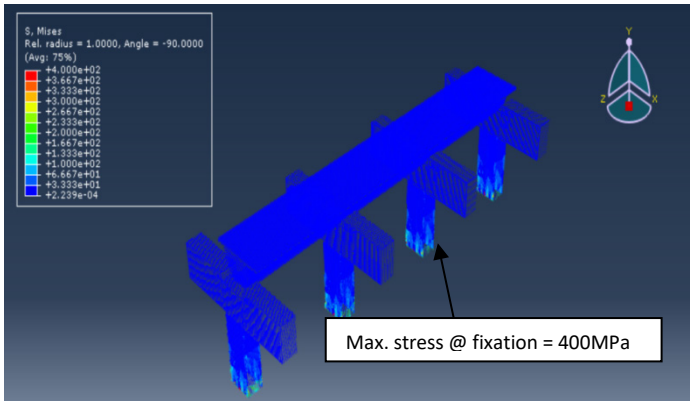


Figure 17: Yield stress for reinforcement = 400MPa

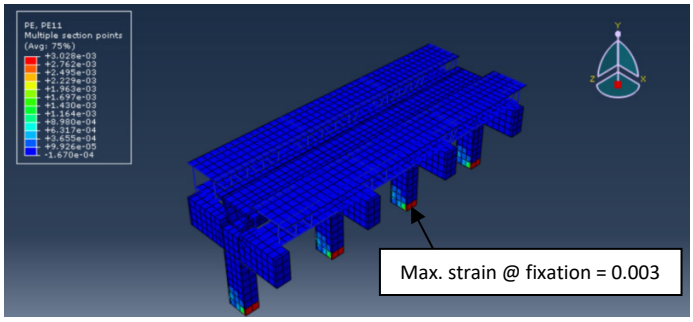


Figure18: Strain for concrete = 0.003

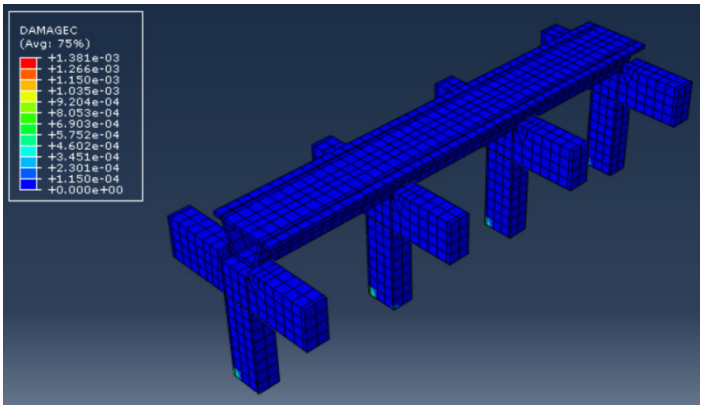


Figure19: Plastic hinges formation & crushing in concrete

9. Observations from The Parametric Study

9.1. Pushover Results

I. 2 Stories:

The results of RC elevated metro stations (2 stories, considering under platform and platform levels) with and without steel cladding are listed in Table 6 and Figure 20.

Table 6: Results for RC (2 stories elevated metro stations) based on pushover analysis

Model	V_u (kN)	V_d[a] (kN)	T[b] (second)	T[c] (second)	R_s[d]	Δ_u[e] (mm)	Δ_y[f] (mm)	μ[g]	R_μ[h]	R[l]
With steel cladding	2392.54	1435.52	0.91	0.52	1.67	14.1	6.8	2.07	1.77	3.47
Without steel cladding	2379.78	1427.86	0.71	0.50	1.67	12.9	6.78	1.88	1.66	3.15

[a]V_d Design base shear equal 60% of the ultimate load capacityas suggested by Uang (1991) [10]

[b]T Calculated time period based on ECP requirements, time period don't exceed:$(1.2*0.075*h^{0.75})$ (5)

[c]T Fundamental period obtained from ETABS model for multi degree of freedom.

[d]$R_s = V_u / V_d$

[e]Δ_u Max Top displacement at V_u, calculated based on peak load, as recommended by Park, R., and Paulay, T,1988 [21].

[f]$\Delta_y$yeild displacement, calculated based on equivalent elasto plastic energy absorption, as recommended by Park, R., and Paulay, T,1988 [21].

[g]μ =Ratio between the ultimate displacement and the yield displacement (Δ_u / Δ_y).

[h]R_μ Function of μ depends on time period:

$$T < 0.2 \text{ seconds } R_\mu = 1$$

$$0.2 < T < 0.5 \text{ seconds } R_\mu = \sqrt{2\mu - 1}$$

$$T > 0.5 \text{ seconds } R_\mu = \mu$$

(1): R calculated based on time period as suggested by Newmark and Hall (1982)[11] in equations (2).

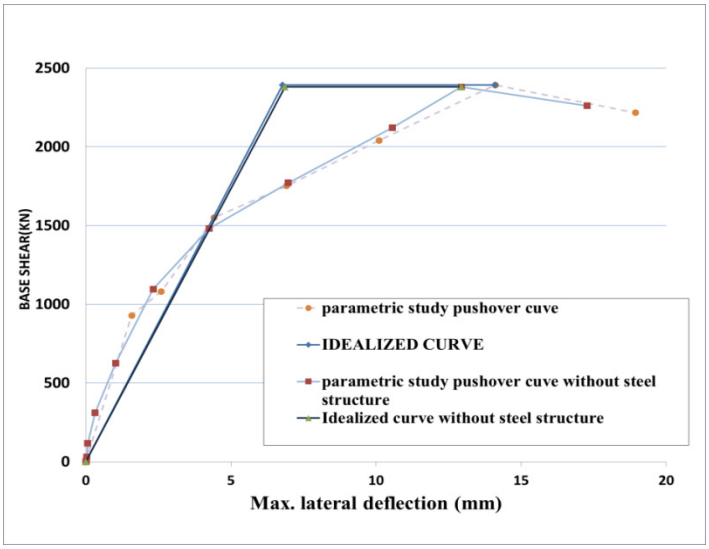


Figure 20: Normalized lateral load versus normalized top displacement

II. 3 Stories:

The results of RC elevated metro stations (3 stories, considering under platform, platform and mechanical levels) with and without steel cladding are listed in Table 7 and Figure 21.

Table 7: Results for RC (3 stories elevated metro stations) based on pushover analysis

Model	V_u (kN)	V_d[a] (kN)	T[b] (second)	T[c] (second)	R_s[d]	Δ_u[e] (mm)	Δ_y[f] (mm)	μ[g]	R_μ[h]	R[l]
With steel cladding	2379.78	1417.88	0.99	0.64	1.67	15.99	7.69	2.08	1.77	3.47
Without steel cladding	2323.24	1393.94	0.79	0.61	1.67	14.1	7.55	1.86	1.65	3.14

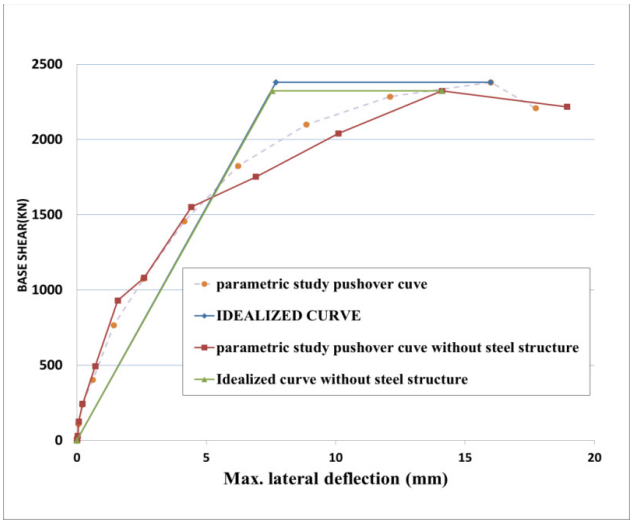


Figure 21: Normalized lateral load versus normalized top displacement

9.2. Time history Results

A series of seven EQs with different frequencies were selected for nonlinear analyses. By reference PEER 2006 database [22], time history of selected earthquakes along with peak ground acceleration (PGA) and year of occurrence of the earthquake are illustrated in Figure 22.

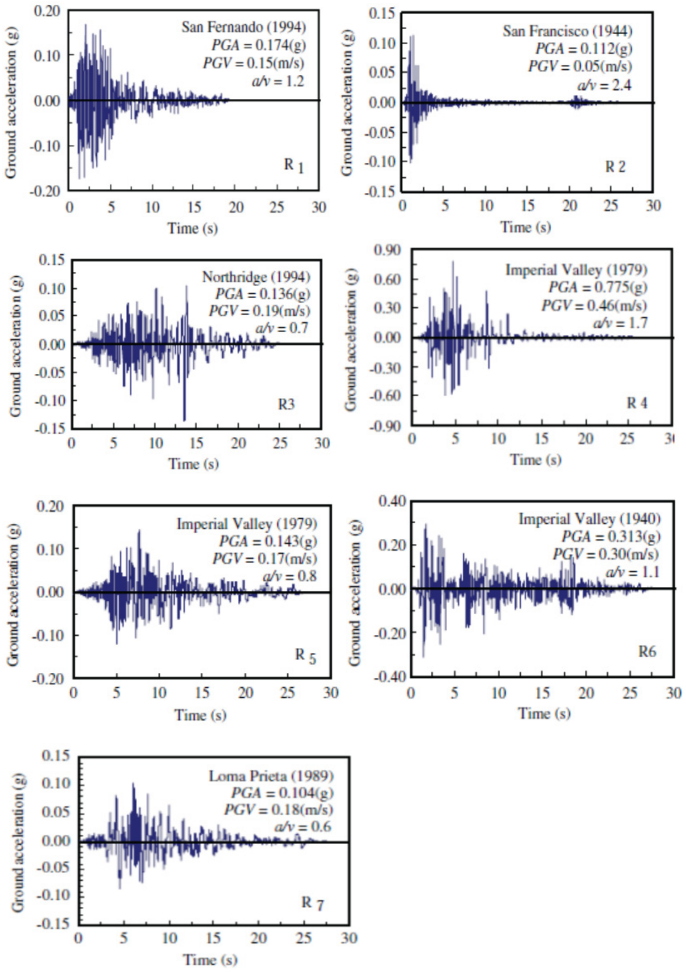


Figure 22: Time history of selected ground motions

Each frame model is subjected to selected records with incremental increase (scaling) of the PGA until a performance limit state is reached. The overall procedure to quantify the seismic force reduction factor (R) is based on the approach presented by Uang (1991) [10]. The R values as shown in Table 9 are assumed as the ratio between elastic lateral load (V_e), which would develop in the seismic force resisting system if the system remained entirely elastic under the design earthquake ground motions, and design lateral force (V_d). A comparison between time history results and pushover results are listed in Table 10.

Table 8: Properties of ground motions (adopted from PEER 2006)

Record	Earthquake	Site	Date	PGA (g)	PGA (MAX) (g)
1	San Fernando	La Hollywood Stor Lot	Feb- 09, 1994	0.174	1.55
2	San Francisco	Golden Gate	March-22, 1944	0.112	2.90
3	Northridge	Playa Del Rey	Jan- 17, 1994	0.136	1.03
4	Imperial Valley	Bonds Corner	Oct- 15, 1979	0.775	1.13
5	Imperial Valley	El-Centro	May- 19, 1940	0.143	1.29
6	Imperial Valley	El-Centro	May- 19, 1940	0.313	1.19
7	Loma Prieta	Apeel Crystal-Spr Res	Oct- 18, 1989	0.104	1.33

Table 9: Results for all reinforced concrete (RC) frames models based on time history analysis

Record	V_d=1417.5 KN	
	V_e (KN)	R= V_e/ V_d
1- San Fernando(1994)	7205	5.03
2- San Francisco(1944)	5668	3.95
3- Northridge(1994)	6498	4.53
4- Imperial Valley(1979)	5300	3.7
5- Imperial Valley(1979)	6756	4.7
6- Imperial Valley(1940)	3805	2.65
7- Loma Prieta(1989)	6164	4.3
	AVG.	3.9

Table 10: Comparison between time history results and pushover results

Time history	Pushover[(a)]	Pushover[(b)]
R	R	R
3.90	3.47	3.17

[(a)]: R calculated based on time period as suggested by Newmark and Hall (1973, 1982) [11] in equation (2), from RC elevated metro station with steel cladding.

[(b)]: R calculated based on time period as suggested by Newmark and Hall (1973, 1982) [11] in equations (2), from RC elevated metro station without steel cladding.

10. Conclusions

This paper is performed to investigate the variation of seismic response of RC structures of different configurations of special RC structures, such as elevated metro stations. Consequently, the response reduction factor can be evaluated. Then the calculated response reduction/modification factors (R) for reinforced concrete (RC) structures will be compared to those specified in ECP [12] and the ASCE code [13].

Some interesting conclusions could be extracted from the parametric study regarding the building behavior, considering that the (R) factor values calculated in this research are related to the studied special structure only, as follows:

1. The response reduction/modification factor(R) value is almost the same as the mentioned value in UBC and ASCE codes. The structural system of the studied RC elevated metro station was considered to be ordinary moment frames according to ACI 318 [23], section 21.2, as columns having a height less than five times the length of the rectangular columns in the moment direction.
2. The (R) value doesn't match those values mentioned in ECP code due to lack of parameters affecting the R value, because the recommended values for (R) factor in the ECP code aren't affected by height of RC structures, which affect ductility factor.
3. The response reduction/modification factor (R) value is sensitive to both RC statically system and RC geometry.
4. The(R) value calculated for elevated metro station with steel cladding is higher than that for elevated metro station without steel cladding, due to increasing of ductility factor.
5. The (R) value decreases with decreasing of RC structure height due to decreasing in ductility factor.
6. The R-value isn't highly affected by addition of mechanical floor because the RC elevated metro station main system is the single RC column (2000x3000) mm, the beam girder (2000x3750) mm and the box girder. And the RC structure stiffness is mainly affected by those elements.
7. The time period calculated based on the structural analysis is less than the calculated time period based on ECP code (equation 5), because ECP code considers the RC structures height only as an affecting parameter on time period. Lesser time period leads to decreasing in (R) factor.
8. The R–values (based on Newmark and Hall assumptions) from pushover analysis are close to those calculated based on time history analysis.

Conflict of Interest

The authors declare no conflict of interest.

Notations

- CDP: Concrete Damaged Plasticity
- EQ : Earthquake
- Es: Steel young's modulus
- F_{cu}: Concrete Compressive Strength
- F_y: Yield stress
- PGA: Peak Ground Acceleration
- PGV:Peak Ground Velocity
- R_μ: Ductility Factor
- R_s: Over-Strength Factor
- RC: Reinforced Concrete
- Rft: Reinforcement
- V_e: Max elastic Base Shear
- V_d: Design Base Shear
- V_u: Maximum Base Shear
- μ : Ductility Capacity

References

[1] Tande. S.N, Ambekar. R.V, "An Investigation of Seismic Response Reduction Factor for Earthquake Resistance Design", International Journal of Latest Trends in Engineering and Technology (IJLTET) DOI: 10.21172 ISSN: 2278-621X Vol.2, Issue 4 July 2013.

[2] Gaurav Rastogi, Khalid Moin and S.M. Abbas, "Non-linear response of a 1:4 scale model of 3-story reinforced concrete frame with partially masonry infill walls", (IJLTET) DOI: 10.21172 ISSN: 2319-8753, Vol.3, Issue 7 July 2014.

[3] Bholebhavi, Inamdar, "AN Evaluation of Response Reduction Factors for irregular structures Using Linear Static Analysis", Issue 5 May 2016. DOI:10.15680/IJIRSET.2016.0505136

[4] Tia Toby, Ajesh K. Kottuppillil," Evaluation of Response Reduction Factor using Nonlinear Analysis", IJIRST: International Journal for Innovative Research in Science & Technology. ISSN (online): 2349-6010, DOI:10.15680, Issue 06 | November 2015.

[5] Minnu M. M, "Evaluation of Response Reduction Factors for Moment Resisting RC Frames", 2014.

[6] Adeel Zafar, "Response Modification Factor of Reinforced Concrete Moment Resisting Frames in Developing Countries", Submitted in partial fulfilment of the requirements for the degree of Master of Science in Civil Engineering in the Graduate College of the University of Illinois at Urbana-Champaign, 2009.

[7] Boumrkik and Kadid, "Pushover Analysis of Reinforced Concrete Frame Structures", Issue 2008.

[8] Applied Technology Council (ATC), Structural response modification factors. Report ATC-19. Redwood City, CA (1995).

[9] Applied Technology Council (ATC). Guidelines for seismic performance Assessment of Buildings (75% Draft): Volume1-Methodology, ATC58-1, Redwood City, CA, 2011.

[10] Uang, C., "Establishing R (or Rw) and Cd factors for building seismic provisions", J. Struct. Eng., 117(1). 19-28(1991). DOI: 10.1061/(ASCE)0733-9445(1991)117:1(19)

[11] Newmark, N. M., and Hall, W. J., "Earthquake spectra and design", EERI Monograph Series, Earthquake Engineering Research Institute, Oakland, Calif, 1982. DOI: 10.1007/s11803-016-0341-1

[12] ECP (Egyptian Code of Practice for Loading),2012.

[13] ASCE 7 (Minimum Design Loads for Buildings and Other Structures. American Society of Civil Engineers. USA), 2005. https://doi.org/10.1061/(ASCE)0733-9445(2005)131:3(507)

[14] UBC (Universal Building Code) 1997. DOI https://doi.org/10.1007/978-1-4615-0481-8_23

[15] ABAQUS software V6.14.

[16] ABAQUS/Standard Analysis User;s Manual, Version 6.14.

[17] Lila M.Abdel-Hafez, "Behavior of masonry strengthened infilled reinforced concrete frames under in-plane load, Issue 21 June 2014. http://dx.doi.org/10.1016/j.hbrcj.2014.06.005

[18] Mansouri, A.," Experimental evaluation of seismic performance of low-shear strength masonry in fills with openings in reinforced concrete frames with deficient seismic details", Issue 11 September 2013. DOI: 10.1002/tal.1115

[19] Aslan Sadeghi Hokmabadi, "Effect of Dynamic soil-pile-structure interaction on seismic response of mid-rise moment resisting framrs",Issue June 2014.

[20] ECP (Egyptian Code of Practice for design), 2017.

[21] Park, R., and Paulay, T "Reinforced Concrete Structures",Wiley,New York, 1988. DOI: 10.1002/9780470172834.ch5

[22] PEER database (PEER 2006).

[23] ACI318m (2014).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 446-450 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Rooftop Urban Agriculture Model with Two Tomato Varieties (*Lycopersicum esculentum* Mill) and Toppings in the High Jungle – Peru

José Alomía-Lucero[1], Jorge Castro-Bedriñana[*,2], Doris Chirinos-Peinado[2]

[1]*Faculty of Agricultural Sciences, Universidad Nacional del Centro del Perú, Satipo, Perú*

[2]*Faculty of Zootechnics, Universidad Nacional del Centro del Perú, Huancayo, Perú*

ARTICLE INFO

Article history:
Received: 20 January, 2020
Accepted: 12 February, 2020
Online: 25 February, 2020

Keywords:
Agroecological system
Food security
Organic tomato
Urban nutrition
Sustainable urban agriculture
Rooftop tomato crops

ABSTRACT

The study objective was to design and evaluate the agroecological system of an agricultural model with vegetable covers (Luffa cylindrica, Cyclanthera pedata and *Mucuna pruriens) associated with tomato plants (Lycopersicum esculentum Mill), Rio Grande and Cherry varieties. The agroecological system is shown with its components of inputs, processes and outputs. The main entrances being solar radiation, rainwater and air as main sources for plant photosynthesis. Components of rooftop agroecosystem are soil, plants, animals and microbes, which form the food pyramid in ecological balance. Production benefits the family in terms of food safety with safe pesticide-free tomatoes. At the same time there is a plant barrier on the roofs to reduce the effects of solar radiation that favors the comfort of the family; in this way the research is framed in the nature of cities with green roofs with their multiple ecological benefits. A new tomato variety obtained by natural crossbreeding between the Cherry and Rio Grande varieties has also been obtained, with medium-sized fruits favorable to the organic tomato market, whose average equatorial diameter is 2.82 ± 0.086 cm, and the average polar diameter 2.47 ± 0.105 cm.*

1. Introduction

The temperature of the atmosphere in the world during the 21st century will continue to increase between 1.0 and 3.7 °C, depending on greenhouse gas emissions [1]; This is expected based on the results obtained from the average of the latest generation of the Earth System Models; so, the global climate will continue in significant warming in response to emissions of CO_2, CH_4 and other greenhouse gases into the atmosphere [2].

The behavior of the maximum temperature and the corresponding thresholds of heat waves have been studied in different places of our planet; In the north-west of the city of Mexicali-Mexico, in recent decades, there is a tendency to increase the temperature, added to a greater number of days with maximum temperatures considered as heat waves [3].

In Perú, the Intergovernmental Panel on Climate Change (IPCC) has collated and reported impacts at the global and regional scales, and global-scale studies concur that Peru is currently highly vulnerable to water stress, with significant variation in the climate across the country.

The climate of the Amazon basin is dominated by the Inter-tropical Convergence Zone (ITCZ), which each year gives very high rainfall (3000 mm per average year). As Peru spans tropical latitudes 0-18 °S, the altitude of the midday sun is never far from vertical and mean monthly temperature is almost constant through the year, particularly so in the cloudy Amazon Basin where the mean temperature is 26 °C throughout the year, rising to 30-32 °C in daytime and falling to 21-22 °C at night [4].

Hot temperature can place many demands on society and can have serious negative impacts over food production and public health. Between variety of methodological approaches, the adaption of agricultural practices to changing climate conditions its included [5], and an adaptation strategy to avoid dangerous levels of climate change would be the use of green roofs.

Climate change has generated many problems, and the intervention of various professions linked to food production in urban environments must be strengthened to improvement the food security of vulnerable populations [6]. The strategy of "greening" the urban space and ensuring the condition of durability of afforestation is necessary to reverse current

[*]Jorge Castro-Bedriñana, Av. Mariscal Castilla N° 3909, El Tambo-Huancayo. +51 964408057. jorgecastrobe@yahoo.com

development trends and their negative effects on the environment and quality of life of people in the urban environment [7].

Urban agriculture is a strategy of environmental sustainability, promoted by green urbanism, and has the ability to meet the multiple demands of food through a considerable diversity of orientations and appropriated technologies [8].

Urban agriculture has great potential to improve the space and life of cities. The installation of orchards and green roofs humanizes the city and becomes more pleasant and comfortable; The humanization of the city involves re-naturalizing it [9]. The naturalizing of roofs with vegetables on light substrates due to the lower weight facilitates their handling on a roof of urban housing [10].

In this context, the objective of the research was to evaluate the components that interact in the design of the system of a model of urban agriculture on roofs with three types of vegetable covers and cultivation of two varieties of tomato (*Lycopersicum esculentum* Mill) in the central jungle of Peru.

2. Methods

2.1. Study place

The investigation was carried out between January and December 2017 in the city of Satipo, located between the geographical coordinates of South Latitude 11° 15′11″ and West Longitude 74° 38′14″, on the eastern foothills of the Andes Mountains, between the Cordillera Oriental and the Amazonian Plain, in the department of Junín. Most of its territory is located in the High jungle and Low jungle ecoregion (dry tropical climate) at an altitude between 207 to 628 m and has a vast biological diversity considered among the richest in the world. It has an area of 81,600 ha.

2.2. Rooftop urban agriculture system design

The dynamic agroecological system design considered:

- Identifying system elements, inputs, processes and outputs.
- Spatial distribution of wooden drawers, with adequate distance for tomato plants per drawer. Three drawers were considered for the Cherry tomato variety and three for the Río Grande variety.
- Design of the model with tomato plants and cover plants.

2.3. Biological material

The urban rooftop farming system considered tomato crops of the Cherry and Rio Grande varieties and cover plants under conditions of central jungle, *Luffa cylindrica*, *Mucuna pruriens* and *Cyclantera pedata*.

3. Results and discussion

3.1. Design of the urban agriculture system of roofs with tomato and vegetable covers

Figure 1 shows the system model with the components of inputs, processes and outputs that are detailed below:

A. Inputs:

Solar energy as a source of energy for photosynthesis, heavy rain as a source of water for plants, and air as a source of oxygen, carbon dioxide and nitrogen.

B. Processes:

- On the soil level: The substrate of agricultural land conditioned in wooden boxes on the roof, in which biological processes of insects and microorganisms for the decomposition of organic matter occur; These allow the biological and chemical fertility of the soil.
- At the producer level: Tomatoes, vegetable covers and weeds, making photosynthesis to transform solar energy into chemical carbohydrate energy accumulating in stems, leaves, fruits and seeds.
- At the first-order consumers level: Hens, arthropods such as pest insects and pest mites, the human family, which feed on the vegetables produced. At the level of microbes: Phyto parasitic nematodes, phytopathogenic fungi, phytopathogenic pseudo fungi, phytopathogenic bacteria and phytopathogenic viruses, symbiotic bacteria of the legume root.
- At the second-order consumers level: Insectivorous and aracnivorous wild birds, hunting spiders, bio controller pests, predatory mites, entomopathogenic fungi, and entomopathogenic bacteria.
- At the decomposer level: Microbes such as facultative fungi and saprophytes, saprophytic bacteria of the soil organic matter.

C. Outputs:

Tomato seeds and fruits, Mucuna and vegetable sponge seeds, free range chicken eggs, and poultry hens for meat.

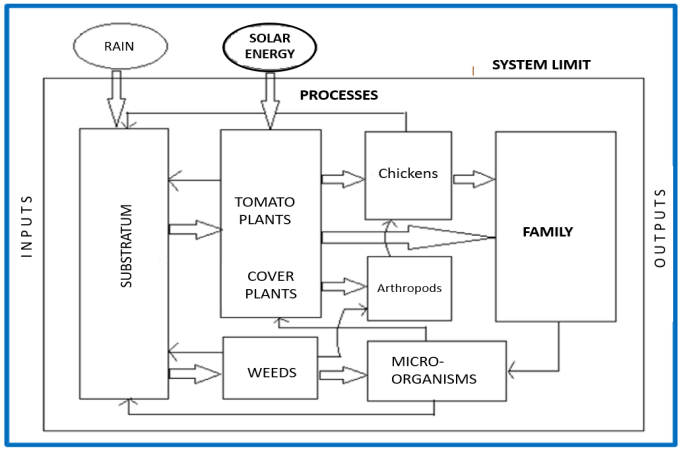


Figure 1. Design of the urban agriculture system of roofs with tomato and vegetable covers

Figure 2 show the spatial distribution of the wooden drawers with the corresponding distance equivalent to 2.70 m^2. Four tomato plants per drawer were considered; The Cherry Tomato variety was grown in three drawers, and the Rio Grande tomato variety in three other drawers; at the ends cover plants were installed, two floors in each drawer.

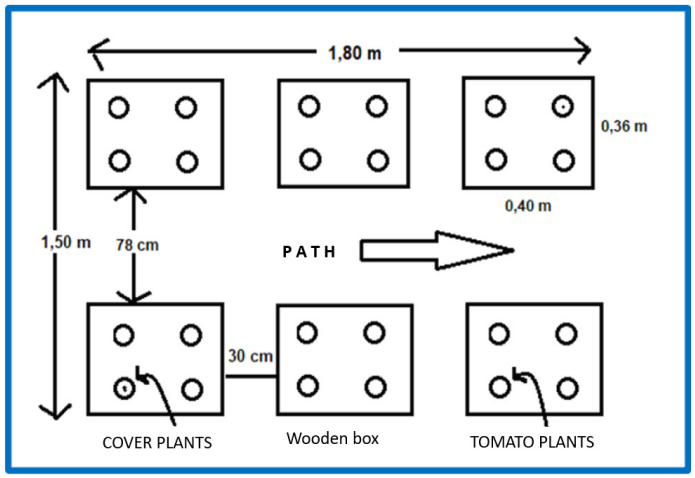


Figure 2. Plan view of the urban rooftop agriculture model with tomato and vegetable covers

Figure 3 show the plan view of the model where tomato plants and cover plants can be seen in the system built with wood. The height of the model generates a volume of 3.51 m^3. The cover plants generate shade to balance the strong hours of sun that support the roofs and the plants.

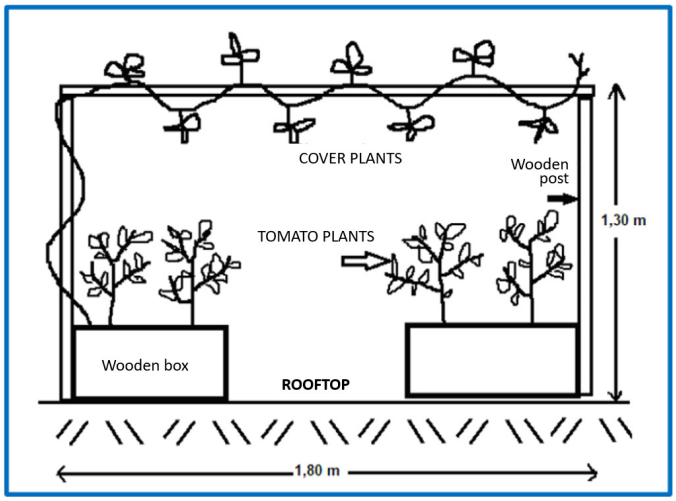


Figure 3. Profile view of the urban agriculture model of roofs with tomato and vegetable covers

Figure 4 shows photographs of the growing *Lycopersicum esculentum* plants, at the beginning with their roof covering plants and ground level.

3.2 Design of the rooftop urban agriculture system model

The Agroecological system inputs are abundant solar energy, rainwater and air with high humidity, which allow the tropical zone to be favorable for plant cultivation. These natural resources allow ecological processes, being at ground level as a substrate of plants, where insects, annelids and microbes that make biological fertility. This low weight substrate has been conditioned in drawers for plants useful growing on the roofs.

At the level of producers, tomato plants, vegetable covers and weeds, perform photosynthesis to transform solar energy into chemical energy which accumulate in the stems, leaves, fruits and seeds. The crops sown and the weeds that arrive by the wind settle in the little substrate and prosper thanks to the abundance of water and sunlight, some are beneficial plants. The important thing about this is that organic tomatoes are harvested that are very healthy [11].

Figure 4. Growing *Lycopersicum esculentum* plants

At the level of first-order consumers, hens feed on the vegetables produced, arthropods and insects [12]. At the microbial level, there are phytoparasitic nematodes, phytopathogenic fungi, phytopathogenic pseudo-fungi, phytopathogenic bacteria and phytopathogenic viruses, and the symbiotic bacteria from legume roots make the system a stable unit.

At the level of second-order consumers, there are insectivorous and aracnivorous wild birds that visit the roofs, hunting spiders, bio controller pests, predatory mites, entomopathogenic fungi and entomopathogenic bacteria. Thus, allowing the ecological balance of the system and to protect biodiversity the use of pesticides for pest control is not necessary [13]. At the decomposing level there are annelids, centipedes, hat fungi, microbes such as facultative fungi and saprophytes, saprophytic bacteria from the soil's organic matter; These organisms allow the system to recycle its organic matter by giving the plants nutrients again.

The outputs are tomato fruits, tomato seeds, mucuna seeds and vegetable sponge, eggs and chicken meat for the benefit of the family, making the system important for food security [8]. Not only vegetables are produced but also animal products [14].

The rooftop agroecosystem has soil, plants, animals and microbes, which form the food pyramid in ecological balance, which allows the sustainability of the system; has higher organic matter, nematode abundances and well tomato fruit yield [15].

The production benefits the family with safe tomatoes, free of pesticides [16]; At the same time there is a plant barrier on the roofs to reduce the effects of solar radiation that harm those who live in the city [17]. In this way the study is framed in the nature of cities with green roofs with ecological benefits [18].

These results demonstrate that roof agriculture is a sustainable alternative, since a place of cement contributes to global warming can now be used for horticultural plants cultivation such as the

tomato Variety Rio Grande, that tolerates high temperatures [19] of the tropical climate of the high jungle.

At soil level the *M. pruriens, C. pedata and L. cylindrica* cover plants find a natural environment suitable for their development and show a high biomass yield, result observed in other study with *M. puriens* [20].

This urban rooftop farming system strengthen food production and food security in cities in a context of vocational training and education in sustainable development, climate change and food, is reinforced [6]. The nature of roofs as one of the alternatives to increase green areas in cities is also urgent as a global proposal [10].

As a complementary result, a new tomato variety was obtained by natural the crossing between the Cherry and Rio Grande varieties. When crossing its stems and floral clusters there was an exchange of pollen from the Rio Grande variety to the Cherry variety. This new strain, which is genetically is F1, has early and late blight tolerance characteristics, with fruits twice as large as Cherry variety, but Rio Grande flavor. This is explained because the tomato is a plant that can be crossed by being close by pollinating insects, which can carry pollen from another plant on their legs and make such pollination.

The equatorial and polar diameters of a sample of 15 tomato cross fruits taken at random from the average were 3.01 ± 0.17 and 2.47 ± 0.26 cm, higher values than the diameters of the Cherry variety, but smaller than the Rio Grande variety.

The equatorial and polar average diameters of a sample of 15 cross tomatoes taken at random were 3.01 ± 0.17 and 2.47 ± 0.26 cm; values greater than the diameters of the Cherry variety, but smaller than the Rio Grande variety. The equatorial diameter is similar to that obtained in the Cherry variety by fertigation, but greater than the polar diameter found by [21], specifying that tomatoes have been organically produced in our case.

A new tomato variety that tolerates heat and blight would be a good alternative for the production of organic tomatoes by reducing inorganic inputs and contributing to the preservation of the environment [22,23]. This study contributes with scientific evidence for food production and consumption in urban areas, and the relationships among city, human well-being and ecosystems [24].

4. Conclusions

The agroecological system designed has with principal inputs the solar radiation, rainwater and air as main sources for plant photosynthesis. The rooftop agroecosystem components are soil, plants, animals and microbes, which form the food pyramid in ecological balance. Production benefits the family in terms of food safety with safe pesticide-free tomatoes. At the same time there is a plant barrier on the roofs to reduce the solar radiation effects that favors the comfort of the family; in this way the work is framed in the nature of cities with green roofs with their multiple ecological benefits.

A new tomato variety obtained by natural crossbreeding between the Cherry and Rio Grande varieties has also been obtained, with medium-sized fruits favorable to the organic tomato market, whose equatorial and polar diameter is 2.82 and 2.47 cm.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

To the families that kindly allowed to install green roofs for the development of the study.

References

[1] IPCC. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. 2013. [OnLine]. Available:https://www.ipcc.ch/site/assets/uploads/2017/09/WG1AR5_Frontmatter_FINAL.pdf.

[2] T. Anderson, E. Hawkins, and P. Jones P. CO_2, the greenhouse effect and global warming: from the pioneering work of Arrhenius and Callendar to today's Earth System Models. Endeavour, vol. 40. no. 3, pp. 178-187. 2016. doi: 10.1016/j.endeavour.2016.07.002.

[3] P. Austria and R. Bandala. Maximum Temperatures and Heat Waves in Mexicali, Mexico: Trends and Threshold Analysis. Air Soil and Water Research, 2016, no. 9, pp. 21-28. doi: 104137-ASWR-S322778.

[4] Met Office. Climate: Observations, projections and impacts: Peru. Department of Energy & Climate Change. Met Office. Devon, EX1 3PB United Kingdom. 2011. [OnLine] Available: http://eprints.nottingham.ac.uk/2040/18/Peru.pdf

[5] J. Luck, M. Spackman, A. Freeman, P. Tre¸bicki, W. Griffiths, K. Finlay and S. Chakraborty. Climate change and diseases of food crops. Plant Pathology, vol. 60, no. 1, pp. 113–121. 2011. doi:10.1111/j.1365-3059.2010.02414.x

[6] M. Chávez y J. Palacios. Importancia de los cursos de agricultura urbana ante el cambio climático y la seguridad alimentaria en México. Grupo eumed.net/Universidad de Málaga y Red Académica Iberoamericana Local-Global. México, vol 10. no 28. 2017. [OnLine]. Available: http://www.eumed.net/rev/delos/28/cursos-agricultura.html

[7] M. Cantón, C. De Rosa y H. Kasperidus. Sustentabilidad del bosque urbano en el área metropolitana de la ciudad de Mendoza. Análisis y diagnóstico de la condición de las arboledas. Avances en Energías Renovables y Medio Ambiente, vol. 7, no 1, pp. 29-34 2003. [OnLine]. Available: https://www.mendoza-conicet.gob.ar/asades/modulos/averma/trabajos/2003/2003-t001-a006.pdf

[8] A. Nadal, E. Cuerva, I. Cerón, A. Josa y J. Rieradevall. Agricultura en edificios urbanos: metodología para la implementación de invernaderos en azoteas de áreas no residenciales urbanas. A: Congreso Internacional Towards Green Cities. "Towards green cities 2016: Compendio de Investigación Académica". Mérida: 2016, p. 429-438. [OnLine]. Available: URI http://hdl.handle.net/2117/100341

[9] R. Marín. La agricultura urbana en el metabolismo de la ciudad: la "burbuja hortícola" de Barcelona y sus elementos de transformación social. XII Congreso Español de Sociología: Grandes Transformaciones Sociales, nuevos desafíos para la sociología. 2016. [OnLine]. Available; http://www.fes-sociologia.com/la-agricultura-urbana-en-el-metabolismo-de-la-ciudad-la-burbuja-hort/congress-papers/2578/

[10] R. Leonardo. Biofertilizantes como opción de naturación de azoteas en zonas urbanas. Universidad Nacional Autónoma de México, Tesis Biología Zaragoza. 76 pp. 2013. [OnLine]. Available: https://www.zaragoza.unam.mx/portal/wpontent/Portal2015/Licenciaturas/biologia/tesis/tesis_leonardo_lopez.pdf

[11] C. Márquez-Hernández, P. Cano-Ríos, Y. I. Chew-Madinaveitia, A. Moreno-Reséndez y N. Rodríguez-Dimas. Sustratos en la producción orgánica de tomate cherry bajo invernadero. Producción orgánica de tomate cherry bajo invernadero. Revista Chapingo Serie Horticultura, vol 12, no. 2, pp. 183-189, 2006.

[12] D. Józefiak and R. Engberg. Insects as poultry feed. Insects in Poultry and Fish Nutrition Conference. 20th European Symposium on Poultry Nutrition Prague, Czech Republic. 2015. [OnLine]. Available: https://www.researchgate.net/publication/281373491_INSECTS_AS_POULTRY_FEED

[13] S. Sepúlveda. Metodología para estimar el nivel de desarrollo sostenible en territorios. Biograma. San José, C.R. IICA 2008. 119 pp. [OnLine]. Available: http://repiica.iica.int/docs/B0664e/B0664e.pdf

[14] Organización de las Naciones Unidas para la Agricultura y la Alimentación - FAO. Agricultura urbana. 2015. [OnLine]. Available: www.fao.org/urban-agriculture/es/

[15] J. Reeves, Z. Cheng, J. Kovach, M. Kleinhenz and P. Grewal. Quantifying soil health and tomato crop productivity in urban community and market gardens. Urban Ecosyst 17, pp 221-238. 2014. https://doi.org/10.1007/s11252-013-0308-1

[16] Azoteas Verdes de Guadalajara. Manual de Agricultura Urbana. pp. 16. 2012. [OnLine]. Available: https://blogdeazoteasverdes.files.wordpress.com/2012/10/manual-agricultura-urbana.pdf

[17] M. Valbuena. Implementación de un modelo de techo verde y su beneficio térmico en un hogar de Honda, Tolima (Colombia) Bogotá. Pontificia Universidad Javeriana. P. 67. 2012. [OnLine]. Available: https://repository.javeriana.edu.co/bitstream/handle/10554/8985/RhodesValbuenaMateode2012.pdf?sequence=1&isAllowed=y

[18] M. H. Zaar. Agricultura urbana: algunas reflexiones sobre su origen y expansión. Biblio 3W. Revista Bibliográfica de Geografía y Ciencias Sociales. Barcelona: Universidad de Barcelona, vol. XVI, no. 944, p. 15. 2011. [OnLine]. Available: http://www.ub.es/geocrit/b3w-944.htm.

[19] J. Santiago, M. Mendoza y F. Borrego. Evaluación de tomate (*Lycopersicon esculentum*, Mill) en invernadero: Criterios fenológicos y fisiológicos. Agronomía Mesoamericana, vol. 9, no. 1, pp. 59-65, 1998. [OnLine]. Available: http://www.mag.go.cr/rev_ meso/v09n01_059.pdf.

[20] Vega P.F. y Leblanc H. Producción de biomasa y fijación de nitrógeno de *Mucuna pruriens* en el trópico húmedo de Costa Rica. Tierra Tropical. vol. 9 no. 1, pp. 57-65, 2013. Available: http://usi.earth.ac.cr/glas/sp/Documentos%202015/Producci%C3%B3n%20de%20biomasa%20y%20fijacion%20de%20nitrogeno.pdf

[21] H. A. Raza et al., "Analysis the effect of 500kv High-Voltage Power Transmission Line on the Output Efficiency of Solar-Panels," in 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 2019, pp. 1–6.

[22] L. Salas, J. González, M. García, E. Sifuentes, S. Parra y P. Preciado. Calidad biofísica y nutracéutica de frutos de tomate producido con sustratos orgánicos. México. Revista Electrónica Nova Scientia. vol. 8, no. 2, pp. 310-325. 2016.

[23] M. Florido y M. Álvarez M. Reseña bibliográfica: Aspectos relacionados con el estrés de calor en tomate (*Solanum lycopersicum* L.) Cultivos Tropicales, vol. 36, no. especial, pp. 77-95. 2015.

[24] F. Orsini, D. Gasperi, L. Marchetti, C. Piovene, S. Draghetti, S. Ramazzotti, G. Bazzocchi and G. Gianquinto. Exploring the production capacity of rooftop gardens (RTGs) in urban agriculture: the potential impact on food and nutrition security, biodiversity and other ecosystem services in the city of Bologna. Food Security, vol. 6, no. 6, pp.781-792. 2014. doi:10.1007/s12571-014-0389-6.