I.J. Wireless and Microwave Technologies, **2021, 5, 18-25**
Published Online October 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2021.05.03

TreeLoc: An Ensemble Learning-based Approach for Range Based Indoor Localization

M.W.P Maduranga
IIC University of Technology, Phnom Penh, 121206, The Kingdom of Cambodia
Email: m.w.pasan@iic.edu.kh

Ruvan Abeysekera
IIC University of Technology, Phnom Penh, 121206, The Kingdom of Cambodia
Email: ruvan@iic.edu.kh

Received: 04 August 2021; Revised: 01 September 2021; Accepted: 18 September 2021; Published: 08 October 2021

Abstract: Learning-based localization plays a significant role in wireless indoor localization problems over deterministic or probabilistic-based methods. Recent works on machine learning-based indoor localization show the high accuracy of predicting over traditional localization methods existing. This paper presents a Received Signal Strength (RSS) based improved localization method called TreeLoc(Tree-Based Localization). This novel method is based on ensemble learning trees. Popular Decision Tree Regressor (DTR), Random Forest Regression (RFR), and Extra Tree Regressor have been investigated to develop the novel TreeLoc method. Out of the tested algorithm, the TreeLoc algorithm showed better performances in position estimation for indoor environments with RMSE 8.79 for the x coordinate and 8.83 for the y coordinate.

Index Terms: Indoor Localization, Machine Learning, Internet of Things, Ensemble Learning, Wireless Sensor Networks.

1. Introduction

Location-Based Services (LBS) are considered primary applications in the Internet of Things (IoT) systems. The recent advancement of IoT has enabled many of the applications in almost every day to day scenario. However, fluctuations and multipath effects, such as Global Positioning Systems (GPS), cannot work effectively in indoor environments such as buildings [1]. Therefore, several types of indoor positioning technologies have been proposed. For example, elderly care, indoor robot navigation applications, ambient assisted living applications could be considered. In all applications above mentions are location-based services. The location-based IoT applications can find the real-time location of a moving human, object, or animal. Many localization techniques and algorithms have been proposed in the literature to find the real-time location of a sensor node in an IoT system. Most of the existing works for indoor localization are based on deterministic algorithms [2-5]. The main drawback of using deterministic or probabilistic algorithms is less efficiency, low prediction accuracy, and difficulty implementing on real IoT devices. In recent contributions to indoor localization, machine learning-based techniques have been introduced as an alternative solution. Many of them are based on supervised learning, and also there are some novel algorithms as well [6-8]. And compare deterministic methods and machine learning methods in [9-10]. The main objective of this work is to develop a novel algorithm with improved accuracy that can be used for indoor localization applications in IoT.

Results in [8-12] show the performance of using tree-based supervised algorithms in localization. Further, Decision Tree Regression (DTR), Extra Tree Regressor (ETR), and Random Forest Regressor (RFR) have been identified as the highly accurate supervised learning algorithms for indoor localization. Therefore, in our work, we intend to optimize tree-based algorithm usage further to build a novel algorithm.

In this work, it has been modified the experimental testbed used in [9]. The testbed consists of three reference nodes and one target node. The main limitations of this testbed are the limited number of references nodes and the environment is affected by multipath fading. In this work, we proposed a novel ensemble learning-based approach to finding the optimal location using a combination of algorithms Decision Tree Regression (DTR), Extra Tree Regressor (ETR), and Random Forest Regressor.

The rest of the paper has been organized as follows. Section II presents the fundamentals of RSS-based localization. Section III presents the experimental setup and dataset used in this work. Section IV presents the data pre-processing techniques applied. Section V presents the development of the novel TreeLoc algorithm, while section VI shows the results.

2. Received Signal Strength (RSS)-Based Indoor Localization.

Received Signal Strength (RSS) is considered a type of signal measurement in indoor localization, and RSS-based measurements have been widely used in many localization systems [6-8]. The relationship between the Received Signal Strength Indicator (RSSI) and the distance is the key to any wireless-based ranging and localization system. RSSI-based position systems use three kinds of propagation models of free space model, bidirectional surface reflection, and log-normal shadowing (LNSM) [9]. As an improved RSSI-based ranging model, we can use the LNSM, which can use for practical applications [11]. It can be defined as;

$$P_r \alpha \frac{p_t}{p^a} \tag{1}$$

where; Pr - Received Signal Pt - Transmit Signal Power d – distance
α – distance power gradient can be found using a table (specific for the environment). According to [8], Eq. 2 can be simplified to get the improved version of LNSM to show the relationship between the RSSI and distance.

$$RSSI = -(10n\log d + A) \tag{2}$$

Where;d – the distance from the mobile/target node to the reference node, n - Signal propagation constant, and A - Received Signal strength at 1m distance.

According to [12-15], trilateration techniques have been used widely to estimate the RSSI-based indoor localization. The time of arrival (TOA) [17] and time difference of arrival (TDOA) [14] are primarily time-based systems that need define with the transmission time. The angle-based arrival angle (AOA) [15] system needs the highly sophisticated directional antenna in beacon nodes to measure angles. Considering the performances of the above range-based techniques used in the literature, RSSI is outperformed. Moreover, techniques such as AoA and TODA need additional hardware to take measurements.

3. Experimental Setup and Data Collection

We used the same testbed in [8] and [9]. This experimental testbed has been set up in an electronic lab with an 8.02 square meter area spanning an open space surrounded by walls. There are three reference nodes placed at the fixed locations and one mobile sensor node. This mobile sensor node considers as our target node to predict the site. The mobile sensor node was kept at 32 different locations from time to time and collected corresponding RSSI values receiving at references nodes.

This experimental testbed is consists of few references nodes and a mobile node developed using microcontrollers. As per fig. 1, RSSI values at the three reference nodes at a particular position of a mobile node will be transmitted to an IoT cloud via a broadband router. The IoT cloud platform is used to integrate the platform with the remote localization model over the internet. Further, the cloud platform is a globally distributed MQTT broker which publishes the information acquired to a remote server. The collected data transfer between the hardware platform and the remote server is accomplished by Wi-Fi and internet technologies, respectively. The ESP8266 has been used to design and develop the mobile and reference nodes in the testbed. ESP8266 is a Wi-Fi module popular for Internet of Things applications [9]. This module has a wireless Wi-Fi transceiver that operates in the IEEE 802.11 b/g/n standard in an unlicensed frequency range of 2400-2484 MHz, supporting the TCP/IP communication protocol stack and Wi-Fi security supporting WAP3.

The system overview of the testbed is shown in figure 3. The RSSI data collected at the mobile sensor node over a private Wi-Fi network is published over the internet, a public network. The three-level architecture used in IoT consists of the sensory, network, and application layers. The sensory layer is responsible for publishing the received signal level to the mobile node. The network layer is the middle layer used as the transporter of the RSSI from mobile to the IoT cloud. The network layer links the RSSI with the remote cloud storage—the application layer used to collect the RSSI data remotely in CloudMQTT. The eclipse plugin is used to remotely collect the RSSI data in the CloudMQTT database for 2D localization [9]. This state-of-art testbed can be implemented for any indoor environment. This architecture provides the feasibility to expand the number of reference nodes. Therefore the size of the testbed is scalable. For example, this testbed could use an indoor application such as elderly activity monitoring, tracking the location of people in a shopping complex, track the location of animals in indoor types farms, etc. Moreover, these sensor nodes are designed on unlicensed 802.11g/802.11b.

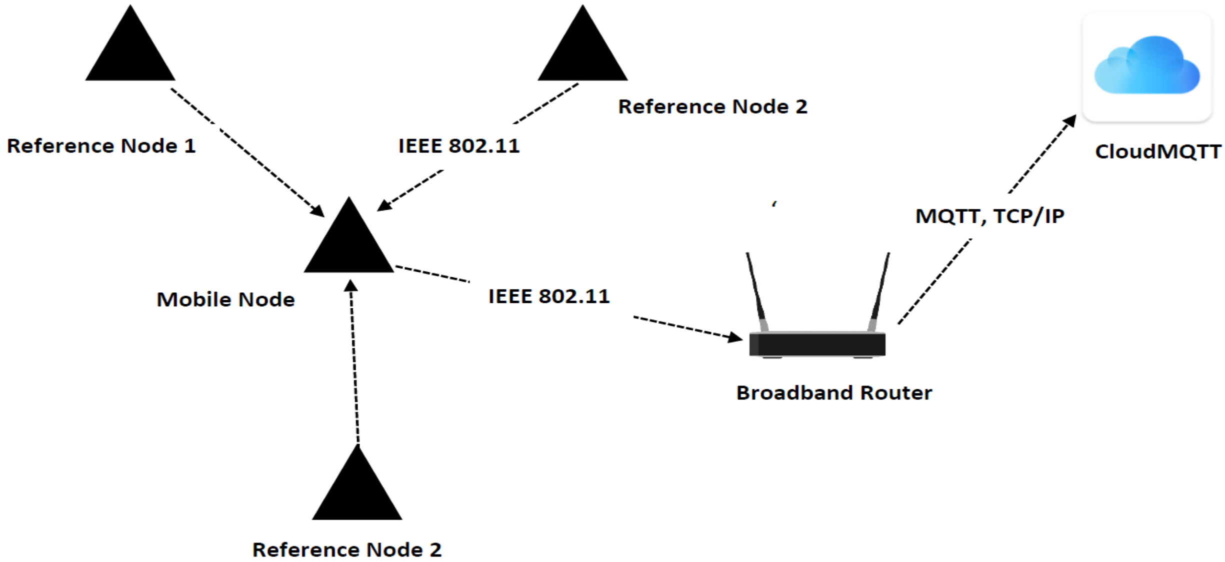

Fig.1. Systems overview of the testbed.

4. Data pre-processing

As RSSI readings are susceptible to noise, data filtering is required. In this experiment, the outliers in the dataset were removed using moving average filters.

Figure 2 shows the raw values of RSSI data collected from three reference nodes for 32 different locations where the RSSI values fluctuate within the range -80 to -20. As per Figure 2, The RSSI's randomness can cause a lot of volatility in the RSSI result. As a result, smoothing the RSSI obtained in real-time is required to lessen the likelihood of localization inaccuracy. The moving average filter can be explained as follows [20].

Let us consider the following dataset.

$$RSSI_0 = [RSSI_0(1), RSSI_0(2), \ldots RSSI_0(n)\ ] \tag{3}$$

Where $RSSI_0$ corresponds to the n^{th} RSSI observed around the target sensor node, now, take the final ten items in 4 and average them as follows:

$$RSSI_1 = [RSSI_1(1), RSSI_1(2), \ldots RSSI_1(n-9)\ ] \tag{4}$$

Where,

$$RSSI_k = \sum_{m=k}^{k+9} RSSI_0\ (m) \qquad and \quad k > 1$$

When the RSSI sample is less than 10, we take the average of all the samples; hence, there will be no delay in real-time position estimation. The result after applying the moving average filter is shown in Figure 3. It is observed that a considerable number of outliers have been removed from the original signal once filtered.

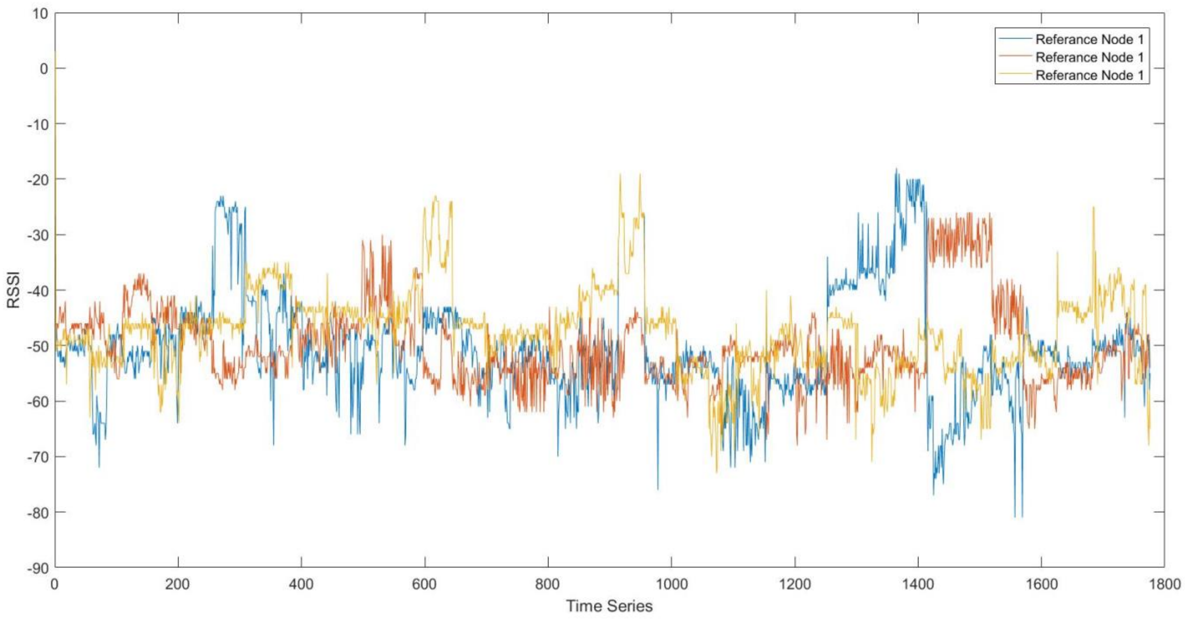

Fig.2. RSSI data without filtering

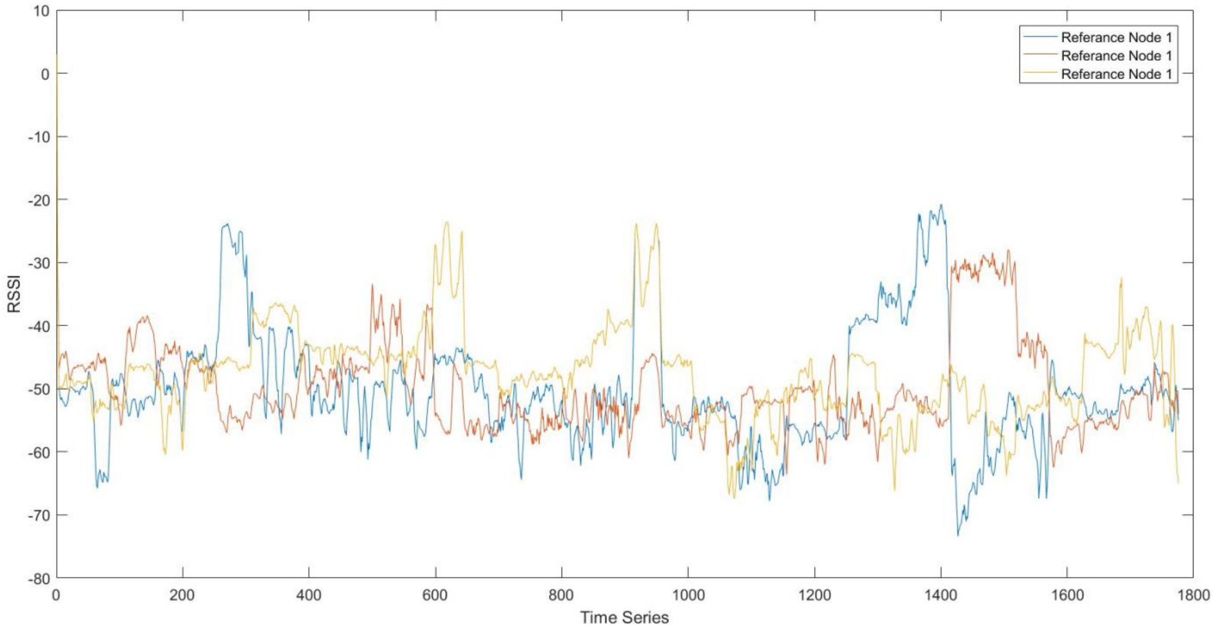

Fig.3. RSSI data after filtering

5. Model Development

5.1. Decision Tree Regression (DTR)

Decision Tree Regressor is a structured type classifier with types of nodes. This structure consists of the root nodes and the interior nodes. The root node represents the whole sample in the dataset, and it splits the entire dataset into sub-nodes. The interior nodes representing the features of a dataset, and decision rules are marked on the branches. The leaf nodes define the outcomes. There could be a linear, non-linear or complex relationship between labels and the features. The number of trees could significantly affect the prediction results. In our model, we have used 40 trees. [17-19].

5.2. Extra Tree Regressor (ETR)

Extremely Randomize Tree Regression (Extra Trees Regression) is a machine learning approach that combines the output predictions of multiple decision trees into a single forecast. It uses a meta estimator that fits several randomized decision trees on various subsamples of the dataset. During the model training, we used ten estimators. Further, it employs a more straightforward approach to create the decision trees used as ensemble members. It can often produce as good as or better performance than the random forest algorithm. It's also simple to use, as it just has a few key hyperparameters and logical heuristics for tuning them. The Extra Trees approach uses the training dataset to generate a massive number of unpruned decision trees. In the case of regression, predictions are made by averaging the forecast of the decision trees [17-19].

5.3.Random Forest Regressor (RFR)

Random Forest Regressor consisting many decision trees, and its consider an ensemble-type machine learning algorithm. A subset of data will be used to train each tree in a random forest. The predictions of each decision tree will be averaged to have the final prediction. Therefor this algorithm is generally powerful than the decision tree algorithm. Few advantages of this algorithm include handling missing values, efficiently handling non-liner paraments, robust outliers of the dataset, and being less impacted by the noises. In our model, we used 100 forests during the simulations [17-19].

5.4. TreeLoc Algorithm

This technique ha a approach on boosting-ensemble method. An ensemble method is a technique that combines the predictions from multiple machine learning algorithms together to make more accurate predictions than any individual model. Few works existing on applying ensemble learning to indoor localization problems [20-23]. In general boosting, it uses a weighted averaging principle to calculate the final prediction [20]. The working of TreeLoc algorithms is shown in figure 6. The full dataset is split into an equal three positions (33.33% each) and fed to the ETR, DTR, and RFR algorithms. The prediction of x and y coordinates from each algorithm is fed into the TreeLoc algorithm. Where TreeLoc will calculate the optimal coordinates again by performing the multiple linear regression. Where coordinates of x and y are generated, each algorithm will use as an independent variable and actual coordinates as the dependent variable to generate the final coordinates. The Eq. x and Eq.x show how final coordinates are calculated.

$$\mathrm{X}_{finally\ predicted} = a + W_1 \times x_{ETR} + W_2 \times x_{DTR} + W_3 \times x_{RFR} \tag{5}$$

$$\mathrm{y}_{finally\ predicted} = b + W_1 \times y_{ETR} + W_2 \times y_{DTR} + W_3 \times y_{RFR} \tag{6}$$

W_1, W_2, and W_3 are the corresponding weighted values for ETR, RFR, and DTR, respectively.

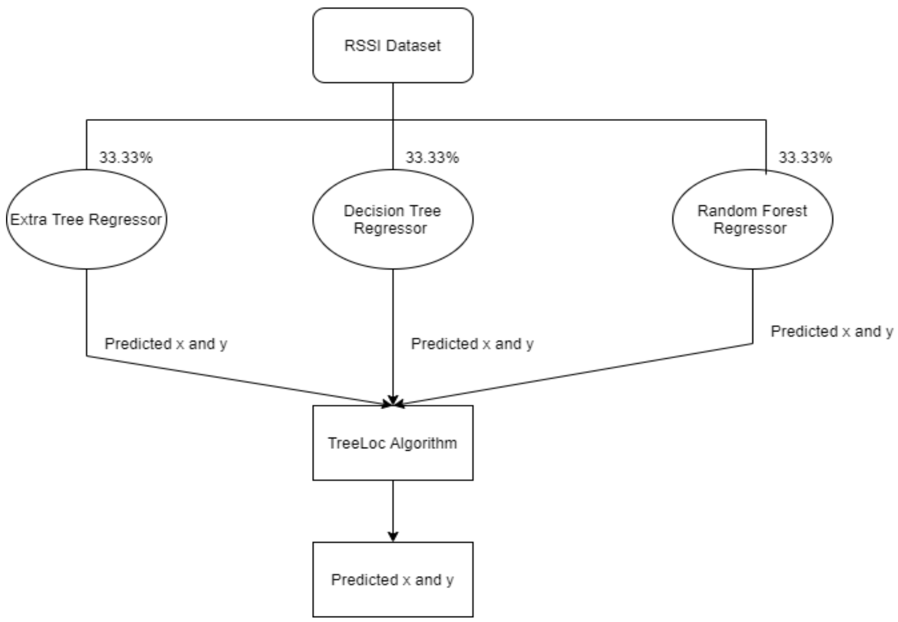

Fig.4. TreeLog algorithm

6. Model Training and Evaluation

Initially, ETR, RFR, and DTR algorithms were computed to observe the performance in location predictions. One thousand seven hundred seventy-six of RSSI data was received from three references nodes used to train the algorithms. The proposed model in section 5 is developed on Jupiter Notebook in python3 environment. And it shows Root Mean Squared (RMSE) and R^2 values as per table 1.

Table 1. Comparison of RMSE and R^2 of algorithms

Algorithm	RMSE		R^2	
	x coordinate	y coordinate	x coordinate	y coordinate
ETR	8.94	9.48	0.9923	0.9184
RFR	29.14	29.09	0.8822	0.9046
DTR	29.33	28.52	0.8952	0.0834
TreeLoc	8.79	8.83	0.9249	0.9244

When considering the prediction results of the above three algorithms, all the algorithms are performed very well in terms of accuracy. All the algorithm shows the error which is less than 30cm. However, when observing the prediction results of all x and y coordinates in individual algorithms mentioned above, each algorithm gives the best prediction in an ad hoc way. Therefore, TreeLoc is introduced to optimize this by taking a weighted average and re-computing the coordinates by performing a multiple linear regression. As per this experiment, obtained models for final x and y are presented in Eq. 7 and 8.

$$\mathrm{X}_{finally\ predicted} = -0.9494 + 0.8036 \times x_{ETR} + 0.5476 \times x_{DTR} + 0.5212 \times x_{RFR} \tag{7}$$

$$\mathrm{y}_{finally\ predicted} = -0.8348 + 0.8922 \times y_{ETR} + 0.5937 \times y_{DTR} + 0.5292 \times y_{RFR} \tag{8}$$

7. Conclusions

In this paper, we propose a novel ensemble-learning-based technique for the indoor localization problem. Popular tree-based algorithms, namely Decision Tree Regression (DTR), Extra Tree Regressor (ETR), and Random Forest Regressor (RFR), are used to build a novel TreeLoc algorithm. Though ETR is outperformed during individual simulations, RFR and DTR also give the best predictions for some samples in an ad-hoc way when observing the prediction results. In the TreeLoc algorithm, we introduced a weighted average mechanism to optimize a location. Simulation results show that the TreeLoc algorithm could provide an 8.79m error for *x* coordinates and an 8.83m error for *y* coordinates.

References

[1] S. Ramnath, A. Javali, B. Narang, P. Mishra, and S. K. Routray, "IoT based localization and tracking," 2017 International Conference on IoT and Application (ICIOT), 2017, pp. 1-4, doi: 10.1109/ICIOTA.2017.8073629.

[2] A. De Blas and D. López-de-Ipiña, "Improving trilateration for indoors localization using BLE beacons," 2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech), 2017, pp. 1-6.

[3] N. S. Kodippili and D. Dias, "Integration of fingerprinting and trilateration techniques for improved indoor localization," 2010 Seventh International Conference on Wireless and Optical Communications Networks - (WOCN), 2010, pp. 1-6, doi: 10.1109/WOCN.2010.5587342.

[4] M. E. Rusli, M. Ali, N. Jamil and M. M. Din, "An Improved Indoor Positioning Algorithm Based on RSSI-Trilateration Technique for Internet of Things (IOT)," 2016 International Conference on Computer and Communication Engineering (ICCCE), 2016, pp. 72-77, doi: 10.1109/ICCCE.2016.

[5] W. Njima, I. Ahriz, R. Zayani, M. Terre and R. Bouallegue, "Smart probabilistic approach with RSSI fingerprinting for indoor localization," 2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2017, pp. 1-6, doi: 10.23919/SOFTCOM.2017.8115509.

[6] A. Gadhgadhi, Y. HachaÏchi and H. Zairi, "A Machine Learning based Indoor Localization," 2020 4th International Conference on Advanced Systems and Emergent Technologies (IC_ASET), 2020, pp. 33-38, doi: 10.1109/IC_ASET49463.2020.9318284.

[7] X. Song et al., "CNNLoc: Deep-Learning Based Indoor Localization with Wi-Fi Fingerprinting," 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2019, pp. 589-595, doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00139.

[8] M W P Maduranga and Ruvan Abeysekara. Supervised Machine Learning for RSSI based Indoor Localization in IoT Applications. International Journal of Computer Applications 183(3):26-32, May 2021.

[9] Y. S. P. Weerasinghe, M. W. P. Maduranga and M. B. Dissanayake, "RSSI and Feed Forward Neural Network (FFNN) Based Indoor Localization in WSN," 2019 National Information Technology Conference (NITC), 2019, pp. 35-40, doi: 10.1109/NITC48475.2019.9114515.

[10] M.W.P Maduranga, Dasuni Ganepola and R.P.S. Kathriarachchi "Comparison of Trilateration and Supervised Learning Techniques for BLE Based Indoor Localization" In Proc. of KDU-14th International Research Conferance,Sri Lanka, 2021

[11] D. Wu, Y. Xu and L. Ma, "Research on RSS based Indoor Location Method," 2009 Pacific-Asia Conference on Knowledge Engineering and Software Engineering, 2009, pp. 205-208, doi: 10.1109/KESE.2009.67.

[12] D. Jia, W. Li, P. Wang and T. Hu, "A range-based localization algorithm for mobile sensor network," 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC), 2017, pp. 599-603, doi: 10.1109/ITOEC.2017.8122367.

[13] M.W.P. Maduranga and A. Taparugssanagorn, RFID Localization Estimation Techniques for Indoor Environment, KDU International Research Conference , Sri Lanka, 2014.

[14] S. R. Jondhale, R. S. Deshpande, S. M. Walke and A. S. Jondhale, "Issues and challenges in RSSI based target localization and tracking in wireless sensor networks," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 2016, pp. 594-598, doi: 10.1109/ICACDOT.2016.7877655.

[15] A. Yassin et al., "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications," in IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 1327-1346, Secondquarter 2017, doi: 10.1109/COMST.2016.2632427.

[16] M. A. Koledoye, D. De Martini, S. Rigoni and T. Facchinetti, "A Comparison of RSSI Filtering Techniques for Range-based Localization," 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), 2018, pp. 761-767, doi: 10.1109/ETFA.2018.8502556.

[17] S. Pathak, I. Mishra and A. Swetapadma, "An Assessment of Decision Tree based Classification and Regression Algorithms," 2018 3rd International Conference on Inventive Computation Technologies (ICICT), 2018, pp. 92-95, doi: 10.1109/ICICT43934.2018.9034296.

[18] A. M. Ahmed, A. Rizaner and A. H. Ulusoy, "A Decision Tree Algorithm Combined with Linear Regression for Data Classification," 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), 2018, pp. 1-5, doi: 10.1109/ICCCEEE.2018.8515759.

[19] X. Wang and Y. Feng, "An Ensemble Learning Algorithm for Indoor Localization," 2018 IEEE 4th International Conference on Computer and Communications (ICCC), 2018, pp. 774-778, doi: 10.1109/CompComm.2018.8780770.

[20] Z. A. Pandangan and M. C. R. Talampas, "Hybrid LoRaWAN Localization using Ensemble Learning," 2020 Global Internet of Things Summit (GIoTS), 2020, pp. 1-6, doi: 10.1109/GIOTS49054.2020.9119520.

[21] D P Gaikwad, "Intrusion Detection System Using Ensemble of Rule Learners and First Search Algorithm as Feature Selectors", International Journal of Computer Network and Information Security(IJCNIS), Vol.13, No.4, pp.26-34, 2021. DOI: 10.5815/ijcnis.2021.04.03

[22] V. Singh, G. Aggarwal and B. V. S. Ujwal, "Ensemble based real-time indoor localization using stray Wi-Fi signal," 2018 IEEE International Conference on Consumer Electronics (ICCE), 2018, pp. 1-5, doi: 10.1109/ICCE.2018.8326317.

[23] N. Akai, T. Hirayama and H. Murase, "Hybrid Localization using Model- and Learning-Based Methods: Fusion of Monte Carlo and E2E Localizations via Importance Sampling," 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020, pp. 6469-6475, doi: 10.1109/ICRA40945.2020.9196568.

[24] Baha Rababah, Rasit Eskicioglu, "Distributed Intelligence Model for IoT Applications Based on Neural Networks", International Journal of Computer Network and Information Security (IJCNIS), Vol.13, No.3, pp.1-14, 2021. DOI: 10.5815/ijcnis.2021.03.01

[25] Shaela Sharmin, Shakil Mahmud Boby, " Characterization of WLAN System for 60 GHz Residential Indoor Environment Based on Statistical Channel Modeling ", International Journal of Wireless and Microwave Technologies (IJWMT), Vol.10, No.2, pp. 42-58, 2020.DOI: 10.5815/ijwmt.2020.02.05

Authors' Profiles

MWP Maduranga obtained his BSc.Eng. in Electronic Engineering degree in 2013 from the Asian Institute of Technology(AIT), Thailand and MS.c.Eng in Electrical and Electronic Engineering degree from the University of Peradeniya, Sri Lanka in 2017. He received Engineering Charter in Electronics and Telecommunication Engineering from the Engineering Council, the UK in 2020. He also a member of IEEE.He is currently a Ph.D. candidate at IIC University of Technology, Cambodia. His current research interests include Machine Learning-based indoor localization, AI/ML in IoT Applications, and wireless communication.

Ruvan Abesekara obtained his Ph.D. in Computer Science and Technology from Dalian Maritime University, China, and holds an MSc in Computer Science from the University of Colombo School of Computing. He is currently the dean of the faculty of Computing, Esoft Metro Campus Sri Lanka, and a researcher at IIC University of Technology, Cambodia. In addition to his academic qualifications, Ruvan holds several industry and professional certifications. He has the membership of several organizations such as The Institute of Doctors Engineers and Scientists, Institution of Engineering and Technology, Institute of Electrical and Electronics Engineers, British Computer Society, Chartered Institute for IT, Computer Society of Sri Lanka, Australian Computer Society, etc. Also, He is a Microsoft Certified Peer Coacher and Cisco Certified Instructor. His current research interests include data mining and algorithms,privacy-preserving, and IoT.

How to cite this paper: M.W.P Maduranga, Ruvan Abeysekera, "TreeLoc: An Ensemble Learning-based Approach for Range Based Indoor Localization", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.11, No.5, pp. 18-25, 2021.DOI: 10.5815/ijwmt.2021.05.03