Empirical Investigation of Multi-tier Ensembles for the Detection of Cardiac Autonomic Neuropathy Using Subsets of the Ewing Features

J. Abawajy[1], A.V. Kelarev[1], A. Stranieri[2], H.F. Jelinek[3]

[1]School of Information Technology, Deakin University
221 Burwood Highway, Burwood, Victoria 3125, Australia
{*jemal.abawajy,kelarev*}*@deakin.edu.au*

[2]School of Science, Information Technology and Engineering
University of Ballarat, P.O. Box 663, Ballarat, Victoria 3353, Australia
a.stranieri@ballarat.edu.au

[3]Centre for Research in Complex Systems and School of Community Health
Charles Sturt University, P.O. Box 789, Albury, NSW 2640, Australia
hjelinek@csu.edu.au

Abstract. This article is devoted to an empirical investigation of performance of several new large multi-tier ensembles for the detection of cardiac autonomic neuropathy (CAN) in diabetes patients using subsets of the Ewing features. We used new data collected by the diabetes screening research initiative (DiScRi) project, which is more than ten times larger than the data set originally used by Ewing in the investigation of CAN. The results show that new multi-tier ensembles achieved better performance compared with the outcomes published in the literature previously. The best accuracy 97.74% of the detection of CAN has been achieved by the novel multi-tier combination of AdaBoost and Bagging, where AdaBoost is used at the top tier and Bagging is used at the middle tier, for the set consisting of the following four Ewing features: the deep breathing heart rate change, the Valsalva manoeuvre heart rate change, the hand grip blood pressure change and the lying to standing blood pressure change.

1 Introduction

Cardiac autonomic neuropathy (CAN) is a condition associated with damage to the autonomic nervous system innervating the heart and highly prevalent in people with diabetes, [6, 7, 24]. The detection of CAN is important for timely treatment, which can lead to an improved well-being of the patients and a reduction in morbidity and mortality associated with cardiac disease in diabetes.

This article is devoted to empirical investigation of the performance of novel large binary multi-tier ensembles in a new application for the detection of cardiac autonomic neuropathy (CAN) in diabetes patients using subsets of the Ewing features. This new construction belongs to the well known general and productive multi-tier approach, considered by the first author in [14, 15].

Standard ensemble classifiers can generate large collections of base classifiers, train them and combine into a common classification system. Here we deal with new large multi-tier ensembles, combining diverse ensemble techniques on two tiers into one scheme, as illustrated in Figure 1. Arrows in the diagram correspond to the generation and training stage of the system, and show that tier 2 ensemble generates and trains tier 1 ensembles and executes them in the same way as it is designed to handle simple base classifiers. In turn, each tier 1 ensemble applies its method to the base classifier in the bottom tier.

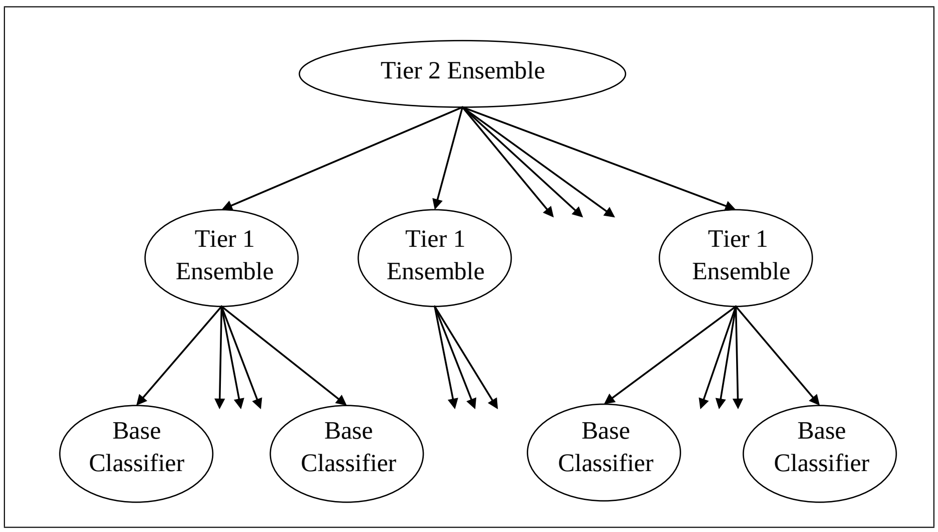


Fig. 1. The generation and training stage of multi-tier ensembles

Large multi-tier ensembles illustrated in Figure 1 have not been considered in the literature before in this form. They can be also regarded as a contribution to the very large and general direction of research devoted to the investigation of various multi-stage and multi-step approaches considered previously by other authors. Let us refer to [1, 14, 15] for examples, discussion and further references.

Our experiments used the Diabetes Screening Complications Research Initiative (DiScRi) data set collected at Charles Sturt University, Albury, Australia. DiScRi is a very large and unique data set containing a comprehensive collection of tests related to CAN. It has previously been considered in [5, 13, 21–23],

For the large DiScRi data set our new multi-tier ensembles produced better outcomes compared with those published in the literature previously. Our new results using multi-tier ensembles achieved substantially higher accuracies.

The paper is organised as follows. Section 2 describes the Diabetes Complications Screening Research Initiative, cardiac autonomic neuropathy and the

Ewing features. Section 3 deals with the base classifiers and standard ensemble classifiers. Section 4 describes our experiments and presents the experimental results comparing the effectiveness of base classifiers, ensemble classifiers and multi-tier ensembles for several subsets of the Ewing features. These outcomes are discussed in Section 5. Main conclusions are presented in Section 6.

2 Diabetes Complications Screening Research Initiative and the Ewing Features

This paper analysed the data set of test results and health-related parameters collected at the Diabetes Complications Screening Research Initiative, DiScRi, organised at Charles Sturt University, [5]. The collection and analysis of data has been approved by the Ethics in Human Research Committee of the university before investigations started. People participating in the project were attracted via advertisements in the media. The participants were instructed not to smoke and refrain from consuming caffeine containing drinks and alcohol for 24 hours preceding the tests as well as to fast from midnight of the previous day until tests were complete. The measurements were recorded in the DiScRi data base along with various other health background data including age, sex and diabetes status, blood pressure (BP), body-mass index (BMI), blood glucose level (BGL), and cholesterol profile. Reported incidents of a heart attack, atrial fibrillation and palpitations were also recorded.

The most essential tests required for the detection of CAN rely on assessing responses in heart rate and blood pressure to various activities, usually consisting of five tests described in [6] and [7]. Blood pressure and heart rate are very important features [2], [29]. The most important set of features recorded for detection of CAN is the *Ewing battery* [6], [7]. There are five Ewing tests in the battery: changes in heart rate associated with lying to standing, deep breathing and valsalva manoeuvre and changes in blood pressure associated with hand grip and lying to standing. In addition features from ten second samples of 12-lead ECG recordings for all participants were extracted from the data base. These included the QRS, PQ, QTc and QTd intervals, heart rate and QRS axis. (QRS width has also been shown to be indicative of CAN [9] and is included here.)

It is often difficult for clinicians to collect all test data. Patients are likely to suffer from other illnesses such as respiratory or cardiovascular dysfunction, obesity or arthritis, making it hard to follow correct procedures for all tests. This is one of the reasons why it particularly important to investigate various subsets of the Ewing battery.

The QRS complex and duration reflects the depolarization of the ventricles of the heart. The time from the beginning of the P wave until the start of the next QRS complex is the PQ interval. The period from the beginning of the QRS complex to the end of the T wave is denoted by QT interval, which if corrected for heart rate becomes the QTc. It represents the so-called refractory period of the heart. The difference of the maximum QT interval and the minimum QT interval over all 12 leads represents the QT dispersion (QTd). It is used as an indicator of the repolarisation of the ventricles. The deflection of the electrical axis of the heart measured in degrees to the right or left is called the QRS axis.

The whole DiScRi database contains over 200 features. We used the following notation for the Ewing features and the QRS width:

$LSHR$ stands for the lying to standing heart rate change;
$DBHR$ is the deep breathing heart rate change;
$VAHR$ is the Valsalva manoeuvre heart rate change;
$HGBP$ is the hand grip blood pressure change;
$LSBP$ is the lying to standing blood pressure change;
QRS is the width of the QRS segment, which is also known as a highly significant indicator of CAN [9].

The detection of CAN deals with a binary classification where all patients are divided into one of two classes: a 'normal' class consisting of patients without CAN, and a 'definite' class of patients with CAN. Detection of CAN allows clinicians to collect fewer tests and can be performed with higher accuracy compared with multi-class classifications of CAN progression following more detailed definitions of CAN progression classes originally introduced by Ewing. More details on various tests for CAN are given in the next section. This paper is devoted to the detection of CAN using subsets of the Ewing features.

A preprocessing system was implemented in Python to automate several expert editing rules that can be used to reduce the number of missing values in the database. These rules were collected during discussions with the experts maintaining the database. Most of them fill in missing entries of slowly changing conditions, like diabetes, on the basis of previous values of these attributes. Preprocessing of data using these rules produced 1299 complete rows with complete values of all fields, which were used for the experimental evaluation of the performance of data mining algorithms.

3 Binary Base Classifiers and Standard Ensemble Methods

Initially, we ran preliminary tests for many binary base classifiers available in Weka [12] and included the following classifiers for a series of complete tests with outcomes presented in Section 4. These robust classifiers were chosen since they represent most essential types of classifiers available in Weka [12] and performed well for our data set in our initial preliminary testing:

- *ADTree* classifier trains an Alternating Decision Tree, as described in [10]. Weka implementation of ADTree could process only binary classes.
- *J48* generates a pruned or unpruned C4.5 decision tree [31].
- *LibSVM* is a library for Support Vector Machines [8]. It can handle only attributes without missing values and only binary classes.
- *NBTree* uses a decision tree with naive Bayes classifiers at the leaves, [28].
- *RandomForest* constructs a forest of random trees following [4].

- *SMO* uses Sequential Minimal Optimization for training a support vector classifier, [19, 30]. Initially, we tested all kernels of SMO available in Weka and used it with polynomial kernel that performed best for our data set.

We used SimpleCLI command line in Weka [12] to investigate the performance of the following ensemble techniques:

- *AdaBoost* training every successive classifier on the instances that turned out more difficult for the preceding classifier [11];
- *Bagging* generating bootstrap samples to train classifiers and amalgamating them via a majority vote, [3];
- *Dagging* dividing the training set into a disjoint stratified samples [33];
- *Grading* labelling base classifiers as correct or wrong [32];
- *MultiBoosting* extending AdaBoost with the wagging [34];
- *Stacking* can be regarded as a generalization of voting, where meta-learner aggregates the outputs of several base classifiers, [35].

We used SimpleCLI command line in Weka [12] to train and test multi-tier ensembles of binary classifiers too.

4 Experimental Results

We used 10-trial 10-fold cross validation to evaluate the effectiveness of classifiers in all experiments. It is often difficult to obtain results for all five tests and we therefore included the largest subsets of four features from the Ewing battery. These subsets can help clinicians to determine whether CAN is present in those situations when one of the tests is missing. The following notation is used to indicate these subsets in the tables with outcomes of our experiments:

S_{Ewing}	is the set of all five Ewing features, i.e., $LSHR$, $DBHR$, $VAHR$, $HGBP$ and $LSBP$;
S_{LSHR}	is the set of four Ewing features with $LSHR$ excluded, i.e., $DBHR$, $VAHR$, $HGBP$ and $LSBP$;
S_{DBHR}	is the set of four Ewing features with $DBHR$ excluded, i.e., $LSHR$, $VAHR$, $HGBP$ and $LSBP$;
S_{VAHR}	is the set of four Ewing features with $VAHR$ excluded, i.e., $LSHR$, $DBHR$, $HGBP$ and $LSBP$;
S_{HGBP}	is the set of four Ewing features with $HGBP$ excluded, i.e., $LSHR$, $DBHR$, $VAHR$ and $LSBP$;
S_{LSBP}	is the set of four Ewing features with $LSBP$ excluded, i.e., $LSHR$, $DBHR$, $VAHR$ and $HGBP$;
S_4	is the set of two heart rate features $LSHR$, $DBHR$, one blood pressure feature $HGBP$, with QRS added.

Feature selection methods are very important, see [25], [26], [27]. In particular, the set S_4 was identified by the authors in [13] using feature selection.

First, we compared the effectiveness of base classifiers for these sets of features. We used accuracy to compare the classifiers, since it is a standard measure of performance. The accuracy of a classifier is the percentage of all patients classified correctly. It can be expressed as the probability that a prediction of the classifier for an individual patient is correct. The experimental results comparing all base classifiers are included in Table 1. These outcomes show that for the DiScRi database RandomForest is the most effective classifier. It is interesting that many classifiers worked more accurately when the LSHR feature had been excluded.

	Subsets of features						
	S_{Ewing}	S_{LSHR}	S_{DBHR}	S_{VAHR}	S_{HGBP}	S_{LSBP}	S_4
ADTree	84.14	84.68	75.31	80.08	81.02	71.73	80.77
J48	91.61	92.15	85.14	90.92	91.28	89.99	91.38
LibSVM	92.39	92.94	80.97	92.71	85.82	84.78	91.13
NBTree	90.15	91.07	81.83	87.45	87.22	86.99	87.76
RandomForest	94.46	94.84	91.76	93.61	94.23	93.76	94.35
SMO	74.13	73.75	64.36	71.98	73.83	71.36	74.44

Table 1. Accuracy of base classifiers for the detection of CAN using subsets of Ewing features

Second, we compared several ensemble classifiers in their ability to improve the results. Preliminary tests demonstrated that ensemble classifiers based on RandomForest were also more effective than the ensembles based on other classifiers. We compared AdaBoost, Bagging, Dagging, Grading, MultiBoost and Stacking based on RandomForest. The accuracies of the resulting ensemble classifiers are presented in Table 2, which shows improvement. We used one and the same base classifier, RandomForest, in all tests included in this table. We tested several other ensembles with different base classifiers, and they turned out worse.

Finally, we compared the results obtained by all multi-tier ensembles combining AdaBoost, Bagging and MultiBoost, since these ensembles produced better accuracies in Table 2. Tier 2 ensemble treats the tier 1 ensemble and executes it in exactly the same way as it handles a base classifier. In turn the tier 1 ensemble applies its method to the base classifier as usual. We do not include repetitions of the same ensemble technique in both tiers, since such repetitions were less effective. The outcomes of the multi-tier ensembles of binary classifiers are collected in Tables 3.

	Subsets of features						
	S_{Ewing}	S_{LSHR}	S_{DBHR}	S_{VAHR}	S_{HGBP}	S_{LSBP}	S_4
AdaBoost	96.84	97.23	94.07	95.99	96.59	96.11	96.51
Bagging	96.37	96.75	93.63	95.52	96.13	95.67	96.05
Dagging	89.75	90.13	87.18	88.94	89.54	89.10	89.46
Grading	94.49	94.87	91.79	93.61	94.26	93.78	94.18
MultiBoost	96.37	96.77	93.62	95.50	96.13	95.65	96.04
Stacking	95.44	95.81	92.70	94.56	95.20	94.73	95.09

Table 2. Accuracy of ensemble classifiers for the detection of CAN using subsets of Ewing features

		Subsets of features						
Tier 2	Tier 1	S_{Ewing}	S_{LSHR}	S_{DBHR}	S_{VAHR}	S_{HGBP}	S_{LSBP}	S_4
AdaBoost	Bagging	97.35	97.74	94.58	96.50	97.12	96.65	97.04
AdaBoost	MultiBoost	96.37	96.78	93.65	95.52	96.14	95.68	96.07
Bagging	AdaBoost	97.33	97.73	94.57	96.49	97.09	96.61	97.00
Bagging	MultiBoost	96.66	97.08	93.91	95.80	96.42	95.96	96.34
MultiBoost	AdaBoost	96.85	97.25	94.08	96.00	96.62	96.13	96.52
MultiBoost	Bagging	97.05	97.43	94.30	96.19	96.83	96.35	96.74

Table 3. Accuracy of multi-tier ensembles of binary classifiers for the detection of CAN using subsets of the Ewing features

5 Discussion

DiScRi is a very large and unique data set containing a comprehensive collection of tests related to CAN. It has been previously considered in [13, 21–23], New results obtained in this paper achieved substantially higher accuracies than the previous outcomes published in [13]. Overall, the results of the present paper are also appropriate for other data mining applications in general when compared to recent outcomes obtained for other data sets using different methods, for example, in [16] and [17].

AdaBoost has produced better outcomes than other ensemble methods for subsets of the Ewing features of the DiScRi data set; and the best outcomes were obtained by a novel combined ensemble classifier where AdaBoost is used after Bagging.

There are several reasons, why other techniques turned out less effective. First, Dagging uses disjoint stratified training sets to create an ensemble, which benefits mainly classifiers of high complexity. Our outcomes demonstrate that the base classifiers considered in this paper are fast enough and this benefit was not essential. Second, stacking and grading use an ensemble classifier to combine the outcomes of base classifiers. These methods are best applied to combine diverse collections of base classifiers. In this setting stacking performed worse than bagging and boosting.

Our experiments show that such large multi-tier ensembles of binary classifiers are in fact fairly easy to use and can also be applied to improve classifications, if diverse ensembles are combined at different tiers. It is an interesting question for future research to investigate multi-tier ensembles for other large datasets.

6 Conclusion

We have investigated the performance of novel multi-tier ensembles for the detection of cardiac autonomic neuropathy (CAN) using subsets of the Ewing features. Our experimental results show that large multi-tier ensembles can be used to increase the accuracy of classifications. They have produced better outcomes compared with previous results published in the literature. The best accuracy 97.74% of the detection of CAN has been achieved by the novel multi-tier combination of AdaBoost and Bagging, where AdaBoost is used at the top tier and Bagging is used at the middle tier, for the set consisting of the following four Ewing features: the deep breathing heart rate change, the Valsalva manoeuvre heart rate change, the hand grip blood pressure change and the lying to standing blood pressure change. This level of accuracy is also quite good in comparison with the outcomes obtained recently for other data sets in closely related areas using different methods, for example, in [18, 20, 16, 17, 36].

Acknowledgements

This work was supported by a Deakin-Ballarat collaboration grant. The authors are grateful to four referees for comments that have helped to improve the presentation, and for suggesting several possible directions for future research.

References

1. Al-Ani, A., Deriche, M.: A new technique for combining classifiers using the Dempster-Shafer theory of evidence. Journal of Artificial Intelligence Research 17, 333–361 (2012)
2. Al-Jaafreh, M., Al-Jumaily, A.A.: New model to estimate mean blood pressure by heart rate with stroke volume changing influence. In: IEEE EMBC 2006: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 1803–1805 (2006)
3. Breiman, L.: Bagging predictors. Machine Learning 24, 123–140 (1996)
4. Breiman, L.: Random Forests. Machine Learning 45, 5–32 (2001)
5. Cornforth, D., Jelinek, H.F.: Automated classification reveals morphological factors associated with dementia. Applied Soft Computing 8, 182–190 (2007)
6. Ewing, D.J., Campbell, J.W., Clarke, B.F.: The natural history of diabetic autonomic neuropathy. Q. J. Med. 49, 95–100 (1980)
7. Ewing, D.J., Martyn, C.N., Young, R.J., Clarke, B.F.: The value of cardiovascular autonomic function tests: 10 years experience in diabetes. Diabetes Care 8, 491–498 (1985)
8. Fan, R.E., Chen, P.H., Lin, C.J.: Working set selection using second order information for training SVM. J. Machine Learning Research 6, 1889–1918 (2005)
9. Fang, Z.Y., Prins, J.B., Marwick, T.H.: Diabetic cardiomyopathy: evidence, mechanisms, and therapeutic implications. Endocr. Rev. 25, 543–567 (2004)
10. Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: Proc. 16th Internat. Conf. Machine Learning. pp. 124–133 (1999)
11. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: Proc. 13th Internat. Conf. Machine Learning. pp. 148–156 (1996)
12. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explorations 11, 10–18 (2009)
13. Huda, S., Jelinek, H.F., Ray, B., Stranieri, A., Yearwood, J.: Exploring novel features and decision rules to identify cardiovascular autonomic neuropathy using a hybrid of wrapper-filter based feature selection. In: Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, ISSNIP 2010. pp. 297–302. IEEE Press, Sydney (2010)
14. Islam, R., Abawajy, J.: A multi-tier phishing detection and filtering approach. Journal of Network and Computer Applications p. to apper soon (2012)
15. Islam, R., Abawajy, J., Warren, M.: Multi-tier phishing email classification with an impact of classifier rescheduling. In: 10th International Symposium on Pervasive Systems, Algorithms, and Networks, ISPAN 2009. pp. 789–793 (2009)
16. Jelinek, H.F., Khandoker, A., Palaniswami, M., McDonald, S.: Heart rate variability and QT dispersion in a cohort of diabetes patients. Computing in Cardiology 37, 613–616 (2010)

17. Jelinek, H.F., Rocha, A., Carvalho, T., Goldenstein, S., Wainer, J.: Machine learning and pattern classification in identification of indigenous retinal pathology. In: 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 5951–5954. IEEE Press (2011)
18. Kang, B., Kelarev, A., Sale, A., Williams, R.: A new model for classifying DNA code inspired by neural networks and FSA. In: Advances in Knowledge Acquisition and Management. Lecture Notes in Computer Science, vol. 4303, pp. 187–198 (2006)
19. Keerthi, S.S., Shevade, S.K., Bhattacharyya, C., Murthy, K.R.K.: Improvements to Platt's SMO algorithm for SVM classifier design. Neural Computation 13(3), 637–649 (2001)
20. Kelarev, A., Kang, B., Steane, D.: Clustering algorithms for ITS sequence data with alignment metrics. In: AI 2006: Advances in Artificial Intelligence, 19th Australian Joint Conference on Artificial Intelligence. Lecture Notes in Artificial Intelligence, vol. 4304, pp. 1027–1031 (2006)
21. Kelarev, A.V., Dazeley, R., Stranieri, A., Yearwood, J.L., Jelinek, H.F.: Detection of CAN by ensemble classifiers based on Ripple Down Rules. In: Pacific Rim Knowledge Acquisition Workshop, PKAW2012. Lecture Notes in Artificial Intelligence, vol. 7457, pp. 147–159 (2012)
22. Kelarev, A.V., Stranieri, A., Yearwood, J.L., Jelinek, H.F.: Empirical investigation of consensus clustering for large ECG data sets. In: 25th International Symposium on Computer Based Medical Systems, CBMS2012. pp. 1–4 (2012)
23. Kelarev, A.V., Stranieri, A., Yearwood, J.L., Jelinek, H.F.: Empirical study of decision trees and ensemble classifiers for monitoring of diabetes patients in pervasive healthcare. In: Network-Based Information Systems, NBiS-2012. pp. 441–446 (2012)
24. Khandoker, A.H., Jelinek, H.F., Palaniswami, M.: Identifying diabetic patients with cardiac autonomic neuropathy by heart rate complexity analysis. BioMedical Engineering OnLine 8, http://www.biomedical–engineering–online.com/content/8/1/3 (2009)
25. Khushaba, R., Al-Jumaily, A., Al-Ani, A.: Evolutionary fuzzy discriminant analysis feature projection technique in myoelectric control. Pattern Recognition Letters 30, 699–707 (2009)
26. Khushaba, R., AlSukker, A., Al-Ani, A., Al-Jumaily, A.: A novel swarm-based feature selection algorithm in multifunction myoelectric control. Journal of Intelligent & Fuzzy Systems 20, 175–185 (2009)
27. Khushaba, R.N., Al-Ani, A., Al-Sukker, A., Al-Jumaily, A.: A combined ant colony and differential evolution feature selection algorithm. In: ANTS2008: Ant Colony Optimization and Swarm Intelligence. Lecture Notes in Computer Science, vol. 5217, pp. 1–12 (2008)
28. Kohavi, R.: Scaling up the accuracy of Naive-Bayes classifiers: a Decision-Tree hybrid. In: Second International Conference on Knowledge Discovery and Data Mining. pp. 202–207 (1996)
29. Mahmood, U., Al-Jumaily, A., Al-Jaafreh, M.: Type-2 fuzzy classification of blood pressure parameters. In: ISSNIP2007: The third International Conference on Intelligent Sensors, Sensor Networks and Information Processing. pp. 595–600 (2007)
30. Platt, J.: Fast training of support vector machines using sequential minimal optimization. In: Advances in Kernel Methods – Support Vector Learning (1998)
31. Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA (1993)

32. Seewald, A.K., Fuernkranz, J.: An evaluation of grading classifiers. In: Advances in Intelligent Data Analysis. Lecture Notes in Computer Science, vol. 2189/2001, pp. 115–124 (2001)
33. Ting, K.M., Witten, I.H.: Stacking bagged and dagged models. In: Fourteenth International Conference on Machine Learning. pp. 367–375 (1997)
34. Webb, G.I.: MultiBoosting: a technique for combining boosting and wagging. Machine Learning 40, 159 – 196 (2000)
35. Wolpert, D.H.: Stacked generalization. Neural Networks 5, 241–259 (1992)
36. Yearwood, J.L., Kang, B.H., Kelarev, A.V.: Experimental investigation of classification algorithms for ITS dataset. In: Pacific Rim Knowledge Acquisition Workshop, PKAW 2008. pp. 262–272. Hanoi, Vietnam, 15–16 December 2008 (2008)

An Uncertainty Quantification Algorithm for Performance Evaluation in Wireless Sensor Network Applications

Dong YU

University of Technology, Sydney, Australia
10743387@student.uts.edu.au

Abstract. The potential applications of Wireless Sensor Networks (WSNs) span a very wide range. One of the application domains is in healthcare industry. The diverse WSN application requirements signify the need of application specific methodology for system design and performance evaluation. Moreover the performance of typical wireless network is stochastic in nature, Probability is an essential instrument needed to assess the performance characteristics. Before WSNs widespread involvement in life or death critical applications, an urgent need is a generic systematic evaluation methodology for decision makers to evaluate performance among alternative solutions taking into account the cohesion characteristic. This paper offers a quantative decision making procedure to incorporate performance deviation as a target performance metric. Decision making is guided by goals and objectives for the particular application specified by application domain experts.

Keywords: Uncertainty Quantification, WSNs, Multi Criteria, Statistical Performance, Generic Methodology, Fair comparison

1 Introduction

We have witnessed in recent years the emergence of WSNs in healthcare. These applications aim to improve and expand the quality of care across wide variety settings and for different segments in the healthcare system. They range from real-time patient monitoring in hospitals, emergency care in large disasters through automatic electronic triage, improving the life quality of the elderly through smart environments, to large-scale field studies of human behavior and chronic diseases [1]. However, the barrier for wide spread adoption of the technology is still high. Fulfilling the potential of WSNs in healthcare requires addressing a multitude of technical challenges. These challenges reach beyond the resource limitations that all WSNs face in terms of limited network capacity and energy budget, processing and memory constraints. Particularly, healthcare applications impose stringent and diverse requirements on system response time, reliability, quality of service, and security.

The uniqueness of WSNs in its resources limitation, transient channel state and drastic different application requirements, brings in application specific system design methodology. From WSNs research kickoff at early stage, research efforts are mostly focused on the isolated programming issue of single layer protocols with little concern of other layer functionality. This leads to protocols that exist in a vacuum which perform well on theoretical basis, but have problems when deployed under real-life circumstances [2, 3]. Many of these protocols are further validated using ad-hoc experimental tests to the benefit of one specific solution, leaving little room for objective comparison with other protocols. Up to now, there exists no fixed set of accepted testing methods, scenarios, parameters or metrics to be applied to guarantee fair comparison between competing solutions. This lack of standardization significantly increases the difficulty for a developer to assess the relative performance of their protocols compared to the current state of the art**.**

Essentially, multiple-criteria evaluation is a well studied realm. There exist plenty of methodologies in multi criteria decision making domain [12]. But we can not apply these techniques directly to WSNs evaluation without considering the uniqueness of the target domain. In this paper, we try to apply analytic hierarchy process (AHP) to WSNs performance evaluation. It is a method to evaluate system performance according to application scenario and application expert preference, it is generic enough to provide a platform to fairly compare alternative solutions. Most importantly, we introduce statistical metrics to reflect uncertainty attribute impact on final performance.

The rest of the paper organized as such: Section 2 establishes a background understanding of performance uncertainty in WSNs that serves as the foundation for building our proposed solutions. The section presents uncertainty attribute of WSN performance, statistical concepts in performance evaluation. Section 3 introduces application specific evaluation based on AHP method. We emphasize that while application specific design is necessary to build efficient application based on limited energy budget, for a fair comparison of alternative design solutions, a generic evaluation algorithm is needed to deal with all of the components aforementioned. Section 4 provides the practical algorithm for uncertainty performance evaluation. Workflow and algorithm are given to get a single QoS performance index. We summarize our work in section 5.

2 Uncertainty Performance in WSNs

2.1 Source of Uncertainty of WSN performance

Several factors contribute to the fact that wireless sensor networks often do not work as expected when deployed in a real-world setting. Firstly, there is possibility of wrong expectation from system designer side: analytical model does not fit into the problem in hands. That is often a problem for inexperienced designers. For all simulation or other experiments methods, first step is to eliminate the possibility of this kind of profound design problem in preliminary stage. Secondly, there is possible wrong expectation from simulation results: Simulation modeling can not faithfully reflect the System Under Test (SUT).

Except the designer's preliminary problem of analytical model mismatching design target, we can further identify fault point of performance disagreement between expectation and real world implementation into components of WSNs hierarchy [11].

1. Environmental influences which may lead to non-deterministic behavior of radio transmission.
2. Node level problem: Malfunction of the sensors, or even the complete failure of a sensor node due to cheap manufacturing cost. Scarce resources and missing protection mechanisms on the sensor nodes may lead to program errors: operating system reliability and fault-tolerance.
3. Network level problem: Network protocols especially (MAC and Routing) are not robust to link failure, contention, topology dynamics.
4. Unforeseen traffic load pattern: A common cause for network problems is an unforeseen high traffic load. Such a traffic burst may occur for example, when a sensor network observes a physical phenomenon, and all nodes in the vicinity will try to report at once, causing the occurrence of packet collisions combined with a high packet loss.

All these factors contribute to the uncertainty of the sensor network behavior and function. These elements increase the probability of network functionality deviation from its normal operation and affects its' collected data accuracy.

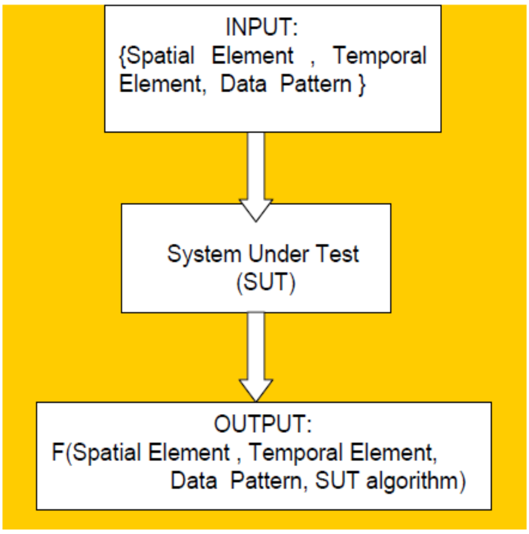


Fig. 1. Uncertainty modeling in WSN

In order to effectively develop parameters, we in [11] congregate hierarchical possible points of deviation into four groups.

1. Spatial elements uncertainty: include site specific characteristic: fading, signal attenuation, interferences, and network scale: topology, network size and density.
2. Temporal elements uncertainty: even on one particular spot, link state flips with time.
3. Data communication pattern uncertainty: include load burst pattern uncertainty (The volume, frequency of the data burst), communication interval difference (how often is the data communication happening, how long is the interval between two adjacent communications), and different communication modes (inquiry triggered, regular report, or event triggered communication).
4. Algorithm internal programming uncertainty: include malfunctioned models, assumption realization problem and other normal cooperation problem in programming,

We summarize above observation into Equation (1).

$$P = F\text{ (spatial, temporal, traffic load pattern, SUT)} \qquad (1)$$

The parameters in equation (1) as show in Figure 1 represent system level performance dynamics involving all four factors. As the input elements display statistical behaviors, output performance definitely will have a statistical distribution pattern with certain norm and deviations for specific scenarios. Since wireless performance is inherently statistical in nature, accurate performance testing must account for these random components [5]. More over, comparing performance curves produced by a number of metrics makes it difficult to evaluate how well a given protocol suits for the purpose of an application. It may also be difficult to estimate, which of the protocols at hand would perform the best with respect to that application [6].

2.2 Characterizing WSNs Uncertainty Performance with Statistical Concepts.

WSNs sense the targeted phenomenon, collect data and make decision to store, aggregate or send the data according to distributed local algorithm. The modulated electromagnetic waves propagate in free-space; interact with the environment through physical phenomenon such as reflection, refraction, fast fading, slow fading, attenuation and human activities. Even with the best wireless protocols, the best chipsets, the best RF design, the best software, wireless performance is going to vary. Wireless performance is inherently statistical in nature, and accurate performance evaluation must reflect this nature.

We observed that currently most ad hoc evaluations in wireless network field, especially in WSNs research, no matter in the form of test bed experiment or simulation, only focus on mean value of performance metrics, and do not pay much attention on performance deviation. For some applications, average performance is sufficient for data gathering and collective data analysis. However, average ‘throughput’, ‘lifetime’, ‘reliability’ or ‘delay response’ are not sufficient enough to predict performance on certain application scenarios. Any dip in performance, no matter how short, can result in dropped packets that cause visual artifacts or pixilation of the image of wireless video monitoring application. In extreme cases like in healthcare emergency monitoring application, any dropped packet may cause life or death difference. Consequently, the viewer/user experience is completely dependent on the wireless system "worst-case" performance [4].

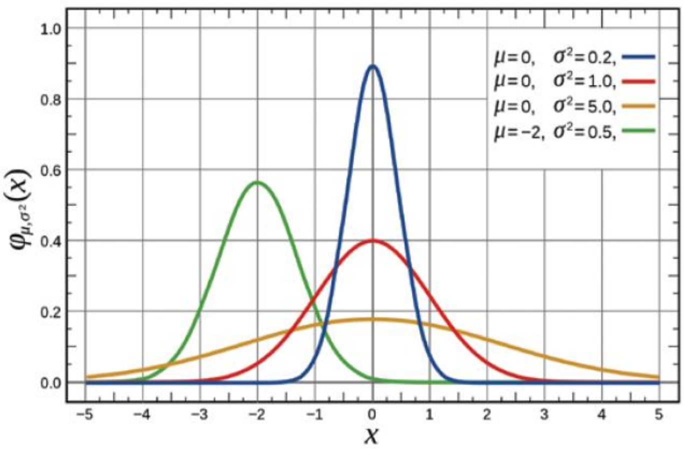


Fig. 2. An example PDF Bell curve graph depicting different average and standard deviation parameters from [5]

Figure 2 represents Probability Density Function" (PDF) of a sample performance metric with four different Systems Under Test (SUT), each with different average and standard deviation (variability) parameters. The graph illustrates „normal" probability distribution revealing statistical characteristics of metric X, representing at least approximately, any variable that tends to cluster around the mean as norm , shown as, but not necessary, the familiar „bell curve". It shows the relative probabilities of getting different values. It answers the question:

- What is the chance I will see a certain result?
- What is the mean value or norm of the respective SUT at this specific performance metric?
- How cohesive and stable is the performance for each SUT?

Examining the random process represented by the red curve in figure 3, we would expect outcomes with a value around „0" to be twice as prevalent as outcomes of around 1.25 (40% versus 20%). However, in some cases, we are more interested in a threshold performance value as benchmark value than individual probability point, what is the probability of having performance being less or greater than a threshold value? A transformed PDF ,Cumulative Distribution Function (CDF) (Figure 3.) helps answer this question.

When you use probability to express your uncertainty, the deterministic side has a probability of one (or zero), while the other end has a flat (all equally probable) probability. For example, if you are certain of the occurrence (or non-occurrence) of an event, you use the probability of one (or zero). If you are uncertain, and would use the expression “I really don't know,” the event may or may not occur with a probability of 50 percent. This is the Bayesian notion that probability assessment is always subjective. That is, the probability always depends upon how much the decision maker knows. Due to statistics science the quality of information and variation are inversely related. That is, larger variation in data implies lower quality data (i.e., information) [7].

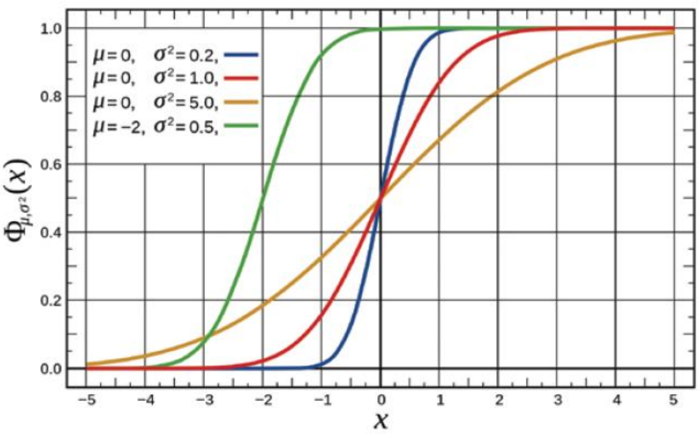


Fig. 3. A typical cumulative distribution function (CDF) graph from [5]

Examining the random performance metric represented by the red curve in figure 3,

- The probability of producing a result less than -0.75 is about 20%.
- The probability of producing a result less than 0 is 50% and
- The probability of producing a result less than 2 is about 95%.

To characterize wireless performance, CDF graphs can provide immensely useful information for system designers to compare performance of alternative solutions. Ideally, we prefer better mean value and smaller deviation, but if the ideal choice is unavailable, we have different optional solution to choose. All the choices should be put in application specific context to choose the right protocol for right application. The principle is:

- Drastic different mean value, if mean value represent positive metrics, like throughput or lifetime, bigger is better, ideally we prefer SUT with bigger mean value and less deviation.
- Same mean value, different deviation: long tail means performance not stable, we prefer smaller deviation.
- Slightly different mean value, but one with long tail, we prefer stable over slightly improved peak performance.
- But if the optimal solution is not available, we have choice over performance stability and higher norm performance according to different application scenario.
  - (option 1) Higher performance potential but less predictable performance
  - (option 2) Less performance potential but higher stable performance

Reference [5] presents practical guidelines on how to actually acquire the statistical performance PDF and CDF curve of a SUT; nevertheless sampling is the key to recover statistical performance and drawing the PDF and CDF curve of a wireless system. Furthermore, to predict real-life performance accurately, researchers ideally should conduct sampling tests across all relevant dimensions and variable if possible. However, in most cases, the design space is too big to exhaustively investigate all factors influencing the final performance. But planners must at least consider three rough dimensions, as we have mentioned above, to characterize wireless performance accurately: time, space, and data pattern. Under each category, there are vast known or unknown parameters that can affect the performance. Hence it is worthwhile to investigate the effect of parameter change to specific performance metric (sensitivity analysis). The effective way to deal with the vast design space is parameter reduction and inter-dependency analysis.

3 Application Specific Evaluation Based on AHP Method

WSNs energy-oriented research originates from conflict between application performance requirements and limited energy supply by battery. In foreseeable future, it will remain as a bottleneck for its widespread development unless a breakthrough at relevant material science field occurs. However, we can not overemphasize energy conservation while ignoring application specific requirements. To what extent we should emphasize the importance of energy aspect comparing other QoS objectives depends on application scenarios. Tradeoffs have to be made on per application basis.

Typically WSN lifetime (energy efficiency), response time (delay), reliability, and throughput are among the main concerns in the design and evaluation process. Under the constraint of wireless sensor node size, the energy budget and computing resources are unfeasible to afford any luxury algorithms. Under such constraint, there does not exist a perfect optimal solution satisfying all performance metrics in the problem (NP hard problem), rather, the question we sought is how to tradeoff multiple criteria explicitly leading to more informed and better decisions. The methodology should be general enough to contain different application scenario according to decision maker's preferences.

There have been few works on application-driven protocol stack evaluation for WSNs. Our evaluation methodology, similar to analytic hierarchy process (AHP) [9, 10], using a Single Performance Index (SPI) for each alternative solution or System Under Test (SUT), as the final quantified goodness measurement for alternative solutions comparison.

The end-user configures the relative importance of the individual design metrics in a function that maps the individual metric values to a single SPI value. Firstly we define the default overall objective function as a weighted sum of the individual design metric normalized values as other AHP methodologies normally do: separation of defining design metric, and weighing those functions' importance in an overall objective function.

$$SPI_{norm} = a * m(L) + b * m(R) + c * m(T); \qquad (2)$$

Here (a, b, c) represent corresponding weight of performance metrics such as lifetime (L), reliability(R) and timeliness (T). (m (L), m (R), m (T)) represents the mean value of multiple measurement of corresponding metric. A key feature of our approach is that, we introduce the statistical analysis of the resulting experiment data, not only using measurement mean value as supposed normalized value (which is not realistic representation of the dynamic truth of the wireless network nature), we introduce deviation of performance measurement PDF as a critical secondary performance metric to emphasis the importance of performance stability and cohesion. Even a higher mean performance metrics, if the performance spreads over a wide spectrum of measurement, not cohesive to so called norm performance (mean) value, it will be problematic for certain application scenarios which require consistent performance , such as health monitoring application and multimedia application. We introduce stability performance index, as:

$$SPI_{stability} = a' * (1/\delta^2(L)) + b' * (1/\delta^2(R)) + c' * (1/\delta^2(T)) \qquad (3)$$

Here (*a', b', c'*) indicates the relative importance of the metrics cohesive characteristic of metrics (L, R, T) represented by deviation ($1/\delta^2$). So overall we have:

$$SPI = SPI_{norm} + SPI_{stability} = \sum_{i=1}^{n} (W_i * Metric_i(mean) + W_i' * (1/\delta_i^2)) \qquad (4)$$

Here 'n' represent the number of metrics considered, W_i = *(a, b, c...)* represents respectively the user specified relative importance of the performance metrics. And W_i' = (*a', b', c'* ...) indicates the relative importance of the metrics cohesive characteristic. The relative importance of each design metric as weight is assigned by considered application specific scenarios, how important in your application is certain metric (network lifetime, reliability, throughput, delay, etc) respectively? How important is cohesive characteristic of performance to your application? Which metric is utmost important for you?

4 The Algorithm Proposed

System evaluation process starts with the end users as application experts who know very well what kind of performance they needs; they specify the most concerned QoS performance metrics, and the weights of each metric.

The WSNs designers decide the initial parameters according to the literature studies and previous experiment experience. Then for each performance metric start the iterative experiment process as such:

1. Parameters significance analysis for $metric_i$.

Repeat *l* experiment measurements, record each experiment the state of each parameter x_i as f_{ji} *f* (1<*i*<*n*, 1< *j*<*l*) and corresponding performance measurement ψ_j (*1<j<l)*. Then use linear aggregation and P-value to decide significant parameters to $metric_i$.

x_1	x_2	$\cdots$	x_n	Ψ
f_{11}	f_{12}	$\cdots$	f_{1n}	Ψ_1
f_{21}	f_{22}	$\cdots$	f_{2n}	Ψ_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
f_{l1}	f_{l2}	$\cdots$	f_{ln}	Ψ_l

2. Design space reduces from *n* parameters to *m* parameters for $metric_i$.

3. *m* parameters interaction analysis for $metric_i$.

Tune the parameters based on the reduced parameters set, Repeat l experiment measurements, record the state of each parameter x_i as f_{ji} (1<*i*<*n*, *1*< *j*<*l*) and corresponding performance measurement ψ_j (*1<j<l)*. Then use the Choquet nonlinear aggregation model as described in later chapter to decide the most effective parameters set including interaction effect of individual parameter.

4. Now we have finally approach the effective parameters set. Tune the effective parameters set, repeat measurement and get the performance curve, get the ($metric_i(\delta^2)$) *and* $metric_i$ (mean) for $metric_i$.

5. Change another metric of interest, start over again from (1).
6. When all metrics of interest finish evaluation, calculate the Single Performance index as aforementioned formula.

$$\text{SPI=}\sum_{i=1}^{n}(W_i * Metric_i(mean) + W_i^{'} * (1/\delta_i^2))$$

7. For competing solution for pair wise comparison, repeat the above process and get SPI value and compare:

```
If system1 SPI > system2 SPI then system1 perform better
                    than system 2
```

Notice that we can setup threshold value as prerequisite filter for minimum requirement, any time if $Metric_i$ mean or deviation is less than the threshold value, the candidate solution is not qualified for further comparison due to unsatisfactory for minimum user specification.

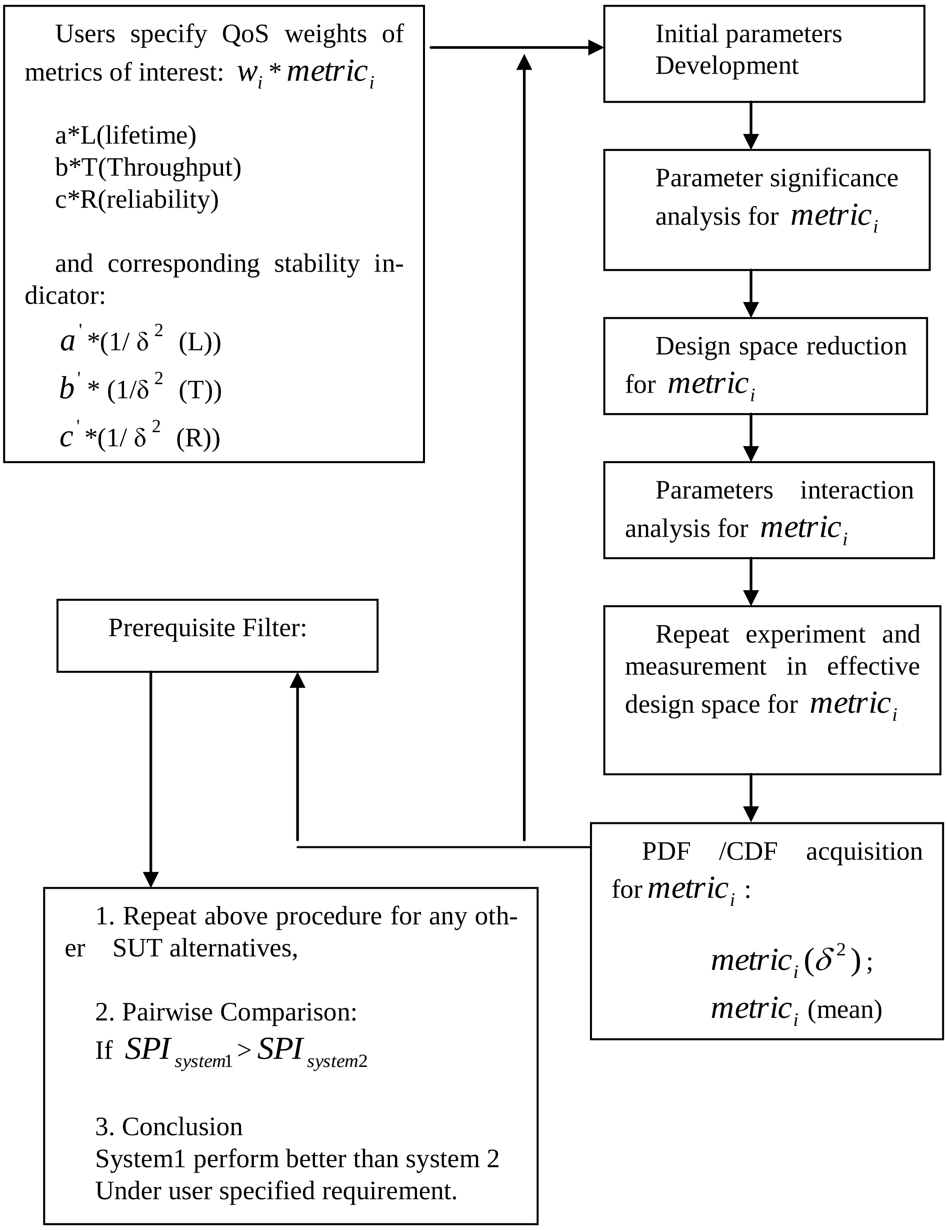
```
Prerequisite Filter:
If  {
every metric_i (mean)> threshold(mean)


And metric_i(δ^2)< threshold(δ^2)
}
Then{
Single Performance Index: SPI=
```

$$\sum_{i=1}^{n}(W_i * Metric_i(mean) + W_i^{'} * (1/\delta_i^2))$$

)

```
Else
SPI=0 (not satisfy minimum requirement,
Not qualified for comparison)
Return
```

Fig. 4. Workflow of proposed benchmarking solution



5 Conclusion and Future Work

In this paper, we introduce a procedure to evaluate WSNs application performance according to application scenarios, in our approach, uncertainty attributes contribute to final performance index. Our future research direction is how to capitalize data mining technique to further dig deep experiment data and distil invaluable collective information from randomness. "Large-scale random phenomena in their collective action create strict, nonrandom regularity" [10].

References

1. Ko, B.J.G.; Lu, C.; Srivastva, M.B.; Stankovic, J.A.; Terzis, A.; Welsh, M. Wireless Sensor Network for Healthcare. *Proc. IEEE* 2010, *98*, 1947-1960.
2. M. Ali, T. Voigt, U. Saif, K. Rmer, A. Dunkels, K. Langendoen, J. Polastre, and Z. A. Uzmi, "Medium access control issues in sensor networks," SIGCOMM 2006, pp. 33–36, 2006.
3. G. Barrenetxea, F. Ingelrest, G. Schaefer, and M. Vetterli, "The Hitchhiker‟s Guide to Successful Wireless Sensor Network Deployments." in SenSys, 2008, pp. 43–56.
4. Yu, D., Nanda, P. He, X. "Performance Uncertainty Impact on WSNs Design Evaluation". IEEE Conference on Control Engineering and Communication Technology (*ICCECT2012*), 7-9 December 2012. Shenyang, China. Accepted.
5. The Ruckus white paper, available at: http://www.ruckuswireless.com/whitepapers/preview/wireless-network-performance
6. J. Haapola, F. Martelli, and C. Pomalaza-R´aez, "Application-driven analytic toolbox for wsns," in 8th International Conference on Ad-Hoc Networks and Wireless (ADHOC-NOW), LNCS 5793, pp. pp. 112 – 125, September 2009.
7. K. Khalili Damghani, M. T. Taghavifard, R. Tavakkoli Moghaddam, "Decision Making Under Uncertain and Risky Situations", Enterprise Risk Management Symposium Monograph Society of Actuaries - Schaumburg, Illinois, 2009.
8. J. Haapola, F. Martelli, and C. Pomalaza-R´aez, "Application-driven analytic toolbox for wsns," in 8th International Conference on Ad-Hoc Networks and Wireless (ADHOC-NOW), LNCS 5793, pp. pp. 112 – 125, September 2009.
9. T. Saaty, *The Analytic Hierarchy Process*. New York: McGraw-Hill, 1980.
10. B. V. Gnedenko and A. N. Kolmogorov *Limit distributions f or sums of independent random variables.* . Trans, and annotated by K. L. Chung. Cambridge, Addison-Wesley, 1954. 264+9 pp.
11. Yu, D., Nanda, P. He, X. "Performance Uncertainty Impact on WSNs Design Evaluation". IEEE Conference on Control Engineering and Communication Technology (*ICCECT2012*), 7-9 December 2012. Shenyang, China
12. Köksalan, M., Wallenius, J., and Zionts, S. (2011). *Multiple Criteria Decision Making: From Early History to the 21st Century*. Singapore: World Scientific

Self-Advising SVM for Sleep Apnea Classification

Yashar Maali[1], Adel Al-Jumaily[1] and Leon Laks[2]

[1]University of Technology, Sydney (UTS)
Faculty of Engineering and IT, Sydney, Australia
Yashar.Maali@student.uts.edu.au, Adel.Al-Jumaily@uts.edu.au
[2] Respiratory Physician, Concord Hospital
Sydney, Australia
leonlaks@bigpond.com

Abstract. In this paper Self-Advising SVM, a new proposed version of SVM, is investigated for sleep apnea classification. Self-Advising SVM tries to transfer more information from training phase to the test phase in compare to the traditional SVM. In this paper Sleep apnea events are classified to central, obstructive or mixed, by using just three signals, airflow, abdominal and thoracic movement, as inputs. Statistical tests show that self-advising SVM performs better than traditional SVM in sleep apnea classification.

Keywords: Sleep apnea, Support vector machines, Particle swarm optimization.

1 Introduction

Sleep disorders are common and sleep apnea (SA) is one of the most common and critical types of sleep disorders. SA can be recognized by the repeated temporary cessation of breathing during sleep [1]. More precisely, apnea is defined as the total or near-total absence of airflow. This becomes significant once the reduction of the breathing signal amplitude is at least around 75% with respect to the normal respiration and occurs for a period of 10 seconds or longer[2]. A sleep apnea event can also be classified into three groups as: central sleep apnea, obstructive sleep apnea, and mixed sleep apnea. In case of the first, sleep apnea is originated by the central nervous system. In the case of the second, the reason for the pauses in the breathing lie in a respiratory tract obstruction, while in the third case, both of these reasons may be present.

The manual scoring of sleep apnea is costly and time-consuming. Therefore, many efforts have been made to develop systems that score the records automatically [11-13]. For this reason several Artificial Intelligent (AI) algorithms are used in this area such as fuzzy rule-based system [14], genetic SVM [15], and PSO-SVM [16] which have been proposed in our previous works. Classification of apneic events to apnea or hypopnea is also so important for severity calculation of the sleep disorder. The classification of apneic events is also considered in many studies, such as [17-19].

In this study an improved version of SVM, named self-advising SVM, is used to classify sleep apnea to central, obstructive and mixed. The second section of this work covers some preliminaries about SVM and partial swarm optimization. We introduce the self-advising SVM algorithm in the third section of this paper; fourth section covers proposed methodology for classifying sleep apnea be the self-advising SVM which is followed by experimental results in section five, and the conclusion in section six.

2 Preliminaries

2.1 Support vector machine

Support vector machine (SVM) is a machine learning method proposed by Vapnik in 1995 [20]. The idea of SVM is to construct a maximized separating hyperplane. The optimization criterion of SVM is the width of the margin between the classes, i.e., the empty area around the decision boundary defined by the distance to the nearest training patterns. SVM shows it ability for classification in many applications even with high dimension. In addition, SVMs avoid over fitting by choosing a specific hyperplane among the many that can separate the data in the feature space.

The brief math description can be shown as follows. For a binary classification, from a training set of N samples$(\boldsymbol{x_1}, y_1), \dots, (\boldsymbol{x_i}, y_i), \dots, (\boldsymbol{x_N}, y_N) \in \mathcal{R}^n \times \{\pm 1\}$, where $\boldsymbol{x_i}$ is the input vector corresponding to the ith sample and labeled by $y_i$depending on its class. SVM aim is, separating the binary labeled training data with a hyperplane that has maximum distance from them, known as maximum margin hyperplane. Figure 1 shows the basic idea of the SVM graphically. The pair $(\boldsymbol{w}, b)$ defines the hyperplane with equation $< \boldsymbol{w}, \boldsymbol{x} > +b = 0$. So, this hyperplane can linearly separate the train data if

$$y_i(\boldsymbol{w}.\boldsymbol{x_i} + b) \geq 1\ , \qquad i = 1, \dots, N. \qquad (1)$$

Distance of each training data $\boldsymbol{x_i}$ from the hyperplane is given by

$$d_i = \frac{\boldsymbol{w}.\boldsymbol{x_i} + b}{\|\boldsymbol{w}\|}, \qquad (2)$$

combining inequality (1) and (2), for all $\boldsymbol{x_i}$ result in

$$y_i d_i \geq \frac{1}{\|\boldsymbol{w}\|}. \qquad (3)$$

Therefore, $\frac{1}{\|\mathrm{w}\|}$ is the lower bound on the distance between the training data $\boldsymbol{x_i}$ and the separating hyperplane.

The maximum margin hyperplane can be considered as the solution of the problem of maximizing the $\frac{1}{\|\mathrm{w}\|}$ subject to the constraint (1), or equivalently by solving the following problem

$$\begin{aligned} &Minimize \quad z = \frac{1}{2}\boldsymbol{w}.\boldsymbol{w} && (4) \\ &s.t. \qquad y_i(\boldsymbol{w}.\boldsymbol{x_i} + b) \geq 1\ , \qquad i = 1, \dots, N. \end{aligned}$$

If we denote $(\alpha_1, \alpha_2, \ldots, \alpha_N)$ the N nonnegative Lagrange multipliers associated with the constraints (1), and without considering few steps the resulting decision function is given by [21],

$$f(\boldsymbol{x}) = sign\left(\sum_{\alpha_i > 0} y_i \alpha_i < \boldsymbol{x}, \boldsymbol{x_i} > +b\right), \tag{5}$$

Note that the nonzero α_i are those for which the constraints (1) are satisfied with the equality sign. This has an important consequence. Since most of the α_i are usually zero the vector $\boldsymbol{w}$ is a linear combination of a relatively small percentage of the training data $\boldsymbol{x_i}$. These points are termed *support vectors* because they are the closest points from the separating hyperplane and the only points needed to determine the hyperplane. The support vectors are the training patterns that lie on the margin boundaries. An advantage of SVM is this fact that only small subset of the training samples, support vectors, is finally retained for the classifier.

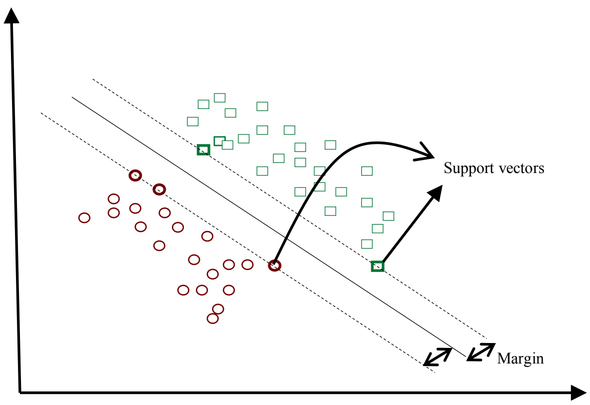


Fig.1. Basic ideas of support vector machines.

In order to use the SVM to produce nonlinear decision functions, the training data is projected to a higher-dimensional inner product space F, called feature space, using a nonlinear map $\phi(\boldsymbol{x}): \mathcal{R}^n \rightarrow \mathcal{R}^d$. In the feature space the optimal linear hyperplane is computed. Nevertheless, by using kernels it is possible to make all the necessary operations in the input space by using $k(\boldsymbol{x_i}, \boldsymbol{x_j}) =< \phi(\boldsymbol{x_i}), \phi(\boldsymbol{x_j}) >$ as $k(\boldsymbol{x_i}, \boldsymbol{x_j})$ is an inner product in the feature space. The decision function can be written in terms of these kernels as follows:

$$f(\boldsymbol{x}) = sign\left(\sum_{\alpha_i > 0} y_i \alpha_i k(\boldsymbol{x}, \boldsymbol{x_i}) + b\right). \tag{6}$$

There are 3 common kernel functions in SVM:

Polynomial kernel : $K(x_i x_j) = (x_i x_j + 1)^q$

RBF kernel : $K(x_i x_j) = e^{-\gamma |x_i - x_j|^2}$

Sigmoid kernel : $K(x_i x_j) = tanh(\gamma x_i^T x_j + c)$

Here q, γ, c are kernel parameters.

2.2 Particle Swarm Optimization

Particle Swarm Optimization (PSO), was introduced by Kennedy and Eberhart in 1995 [22, 23] based on the movement of swarms and inspired by the social behaviors of birds or fishes. Similar to the genetic algorithm, PSO is a population-based stochastic optimization technique. In the PSO, each member is named particle, and each particle is a potential solution to the problem. In comparison with the genetic algorithms, PSO updates the population of particles by considering their internal velocity and position, which are obtained by the experience of all the particles.

In this study, we use the constriction coefficient PSO [24]. In this approach, the velocity update equation is as (1),

$$v_{ij}(t+1) = \chi \left[v_{ij}(t) + \phi_1 \left(y_{ij}(t) - x_{ij}(t)\right) + \phi_2 \left(\hat{y}_{ij}(t) - x_{ij}(t)\right)\right], \quad (7)$$

where y_{ij} is the particle best and $\hat{y}_{ij}$ is the global best particles. And,

$$\chi = \frac{2k}{|2 - \phi - \sqrt{\phi(\phi - 4)}|}, \quad (8)$$

with,

$$\phi = \phi_1 + \phi_2, \quad \phi_i = c_i r_i \quad i = 1,2.$$

Equation (7) is used under the constraints that $\phi \geq 4$ and $k, r_i \in [0,1]$.

The parameter k in the equation (8) controls the exploration and exploitation. For $k \sim 0$, fast convergence is expected and for $k \sim 1$ we can expect slow convergence with a high degree of exploration [24].

The constriction approach has several advantages over traditional PSO model such as; we do not need velocity clamping for constriction model and this model guarantees convergence under the given constraints[25].

3 Self-Advising SVM

In current SVM methods, the only information that is used in the test phase from the training is the hyperplane positions or SVs. Subsequent knowledge can be any more information about the SVs, such as their distribution, and or the knowledge extracted from the misclassified data in the training phase.

Self-advising SVM tries to generate subsequent knowledge from the misclassified data of the training phase of the SVM. This misclassified data can come from 2 potential sources as outliers or as data that have not been linearly separated by using any type of kernels. Classic SVM ignores the training data that has not been separated linearly by kernels in the training phase. Self-advising SVM intended to deal with the ignoring of the knowledge that can be extracted from the misclassified data. This can be done by generating advice weights based on using of misclassified training data, if possible, and use these weights together with decision values of the SVM in the test phase. These weights help the algorithm to eliminate the outlier data.

To benefit from the misclassified data of the training phase, we must first find them. Let's define the misclassified data sets, MD, in the training phase as follows:

$$MD = \bigcup_{i=1}^{n} \boldsymbol{x}_i | y_i \neq sign\left(\sum_{\alpha_j > 0} y_j \alpha_j k(\boldsymbol{x_i}, \boldsymbol{x_j}) + b\right) \tag{9}$$

It must be considered that on the right hand side of the equation (9), we can use any SVM decision function and kernel. The MD set can be null, but experimental results revealed that the occurrence of misclassified data in training phase is common.

For each $\boldsymbol{x}_i$ of MD the neighborhood length (NL) is defined as:

$$NL(\boldsymbol{x}_i) = minimum_{\boldsymbol{x}_j}\left(||\boldsymbol{x}_i - \boldsymbol{x}_j|| \middle| y_i \neq y_j\right). \tag{10}$$

where $\boldsymbol{x_j}$, $j = 1, \dots, n$, are the training data.

Note: if the training data is mapped to a higher dimension by using a mapping function, then the distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be computed according to the following equation with reference to the related kernel k,

$$||\boldsymbol{\theta}(\boldsymbol{x}_i) - \boldsymbol{\theta}(\boldsymbol{x}_j)|| = \left(k(\boldsymbol{x}_i, \boldsymbol{x}_i) + k(\boldsymbol{x}_j, \boldsymbol{x}_j) - 2k(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)^{0.5}. \tag{11}$$

Finally, based on finding of NL, for each $\boldsymbol{x}_k$ from the test set, the advised weight (AW) is computed as follows,

$$\begin{cases} 0, & \forall \boldsymbol{x}_i \in MD,\ \left||\boldsymbol{x}_k - \boldsymbol{x}_i|\right| > NL(\boldsymbol{x}_i)\ or\ MD = NUL \\ 1 - \frac{\sum_{x_i} ||x_k - x_i||}{\sum_{x_i} NL(x_i)} & \boldsymbol{x}_i \in MD,\ \left||\boldsymbol{x}_k - \boldsymbol{x}_i|\right| \leq NL(\boldsymbol{x}_i) \end{cases} \tag{12}$$

These AWs are between 0 and 1, and they represent how close the test data are to the misclassified data. To conclude the above, the self-advising SVM (SA-SVM) is as follows:

Training phase:

1- Finding the hyperplane by solving problem of equation (6) or related problem, it means normal SVM training.
2- Find the MD set using equation (9).
3- If the MD is null, go to the testing phase else compute NL for each member of MD using equation (10).

Testing phase:

1- Compute the $AW(x_k)$ for each x_k from the test set

2- Compute the absolute value of the SVM decision values for each x_k from the test set and scale the values to $[0,1]$.

3- For each x_k from the test set,

If $AW(x_k) < \text{desicion value } (x_k)$ then $y_k = \text{sign}\left(\sum_{\alpha_j>0} y_j \alpha_j k(x_k, x_j) + b\right)$, this means normal SVM labeling.

Else, $y_k = y_i| \, (||x_k - x_i|| \leq NL(x_i) \text{ and } x_i \in MD)$.

Note: If the testing and training data are mapped to a higher dimension, then $||\boldsymbol{x}_k - \boldsymbol{x}_i||$ in step 3 of the test phase should be computed by equation (11); further, as mentioned previously, any SVM methods and kernels can be used in this algorithm.

4 Approach and Method

In this section, we present the proposed algorithm for the classification of the sleep apnea events into central, obstructive or mixed. The proposed methodology is as follows:

- Feature generation: this stage generates several statistical features for each event from the wavelet packet coefficients.
- PSO-SVM classifier: In this stage PSO is used to select a best features subset interactively with the SA SVM. PSO also is used for tuning the parameters of the SA SVM. In the process, SA SVM is used as the fitness evaluator.
- Final classification: the selected pattern is used for classification of the unseen validation data in this stage. The accuracy of this step is assumed as the final performance of the algorithm.

The details of these steps are as follows:

4.1 Feature generation

Feature extraction plays an important role in recognition systems, in this paper features are generated from wavelet coefficients. In the first step, 3 levels "Haar" wavelet packet applied on input signals, airflow, abdominal and thoracic movements. Then several statistical measures are computed by attention to the coefficients related to each apnea events, and considered as features of that event. These features represent the inputs of proposed PSO-SVM algorithm in the next step. Full list of proposed features are included in Table I.

Table I: List of statistical features, x is coefficients of wavelet.

$\log(\text{mean}(x^2))$	$\text{kurtosis}(x^2)$	$\text{geomean}(\lvert x\rvert)$
$\text{std}(x^2)$	$\text{var}(x^2)$	$\text{mad}(x)$
$\text{skewness}(x^2)$	$\text{mean}(\lvert x\rvert)$	$\text{mean}(x^2)$
$\text{skewness}(x)$	$\text{kurtosis}(x)$	$\text{var}(x)$
$\text{geomean}(x^2)$	$\text{mad}(x^2)$	$\text{std}(x)$

4.2 Particle representation

In this study, each particle consists of two arrays; the length of the first array is equal to the number of features. Each cell can get a real number between 0 and 1 as importance of the relevant feature. Features, which their corresponding cells have values higher than 0.5, are selected for classification. The second array is related to the gamma and cost as parameters of the SVM, which can get a value between 2^{-5} to 2^{5}.

5 Results and discussion

Experimental data consist of 20 samples which events of them are annotated by an expert were provided by the Concord hospital in Sydney. We run the algorithm 5 different times; in each run 10 samples are chosen as the training, 5 samples as validation and 5 samples as the test set. RBF kernel is selected for the both of the Self-advising and traditional SVM. In the constriction coefficient PSO structure, k considered as 0.8 and $c_1 = 2, c_2 = 4$ and swarm contain 20 particles

Table 2 tabulates the number of central, obstructive and mixed events in each of the validation set, train and test for these 5 runs.

Accuracy and also f-score of the self-advising and traditional SVM in classification of these apnea events are as Table 3.

Table 2. Number of obstructive, central and mixed apnea in 5 different runs

	Obstructive	central	Mixed
#1	931	375	312
#2	879	463	276
#3	913	478	227
#4	870	494	254
#5	894	453	271

Table 3. Accuracies and f-score of self-advising and traditional SVM in classification of apnea events

	SVM		Self-Advising SVM	
	Accuracy	f-score	Accuracy	f-score
#1	85.02	0.79	87.32	0.84
#2	86.31	0.83	86.81	0.84
#3	78.44	0.74	83.28	0.81
#4	77.45	0.79	79.95	0.80
#5	76.44	0.77	82.29	0.84
Total	80.732	0.784	83.93	0.826

The average accuracies for these two methods are as 80.73 and 83.93, respectively. Also, for more reliable evaluation between results of these two methods pair t-test is used. The p value of t-test is as 0.028. These statistical tests show that the results obtained by the self-advising SVM are significantly better than the results of traditional SVM.

Also, we consider the f-score as another performance measure to compare these methods. Table 3, tabulated the f-scores for these two methods. The average f-score for traditional SVM and self-advisable SVM are 0.78 and 0.83, repectively. Also paired t-test shows that self-advisable is significantly better than traditional SVM by considering the f-score. The p value of t-test is as for the f-score is as 0.036.

6 Conclusion

In this study we proposed a new version of SVM named self-advising SVM for classification of sleep apnea events into obstructive, central or mixed. This study shows that self-advising SVM has advantage over traditional SVM in apnea classification problem. More investigation of the proposed SVM algorithm in apnea detection or other classification problems must be study in future works.

References

1. *Guilleminault, C., J. van den Hoed, and M. Mitler, Overview of the sleep apnea syndromes. In: C. Guilleminault and Wc Dement, Editors, Sleep apnea syndromes, Alan R Liss, New York. 1978: p. 1-12.*
2. *Flemons, W.W.; Buysse, D; Redline, S, Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. Sleep, 1999. 22(5): p. 667-689.*
3. *Chokroverty, S., Sleep deprivation and sleepiness, in Sleep Disorders Medicine (Third Edition)2009, W.B. Saunders: Philadelphia. p. 22-28.*
4. *Chokroverty, S., Overview of sleep & sleep disorders. Indian Journal of Medical Research, 2010. 131(2): p. 126-140.*
5. *Ball, E.M.; Simon RD Jr; Tall AA; Banks MB; Nino-Murcia G; Dement WC, Diagnosis and treatment of sleep apnea within the community - The Walla Walla project. Archives of Internal Medicine, 1997. 157(4): p. 419-424.*
6. *Kryger, M.H.; Roos L, Delaive K, Walld R, Horrocks J, Utilization of health care services in patients with severe obstructive sleep apnea. Sleep, 1996. 19(9): p. S111-S116.*
7. *Stradling, J.R. and J.H. Crosby, Relation between systemic hypertension and sleep hypoxemia or snoring- analysis in 748 men drawn from general-practice. British Medical Journal, 1990. 300(6717): p. 75-78.*
8. *Hoffstein, V., Snoring, in Principles and practice of sleep medicine, M.H. Kryger, T. Roth, and D.W. C., Editors. 2000, Saunders: Philadelphia, PA: W.B. p. 813-826.*
9. *Penzel, T.; McNames J; Murray A; de Chazal P; Moody G; Raymond B., Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. Medical & Biological Engineering & Computing, 2002. 40(4): p. 402-407.*
10. *Kryger, M.H., management of obstractive sleep-apnea. Clinics in Chest Medicine, 1992. 13(3): p. 481-492.*
11. *Cabrero-Canosa, M., E. Hernandez-Pereira, and V. Moret-Bonillo, Intelligent diagnosis of sleep apnea syndrome. Ieee Engineering in Medicine and Biology Magazine, 2004. 23(2): p. 72-81.*
12. *Cabrero-Canosa, M; Castro-Pereiro, M; Graña-Ramos, M; Hernandez-Pereira, E; Moret-Bonillo, V; Martin-Egaña, M; Verea-Hernando, H, An intelligent system for the detection and interpretation of sleep apneas. Expert Systems with Applications, 2003. 24(4): p. 335-349.*
13. *de Chazal, P.; Heneghan C; Sheridan E; Reilly R; Nolan P; O'Malley M, Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. Ieee Transactions on Biomedical Engineering, 2003. 50(6): p. 686-696.*
14. *Maali, Y. and A. Al-Jumaily, Genetic Fuzzy Approach for detecting Sleep Apnea/Hypopnea Syndrome. 2011 3rd International Conference on Machine Learning and Computing (ICMLC 2011), 2011.*

15. *Maali, Y. and A. Al-Jumaily. Automated detecting sleep apnea syndrome: A novel system based on genetic SVM. in Hybrid Intelligent Systems (HIS), 2011 11th International Conference on. 2011.*
16. *Maali Y. and A. Al-Jumaily, A Novel Partially Connected Cooperative Parallel PSO-SVM Algorithm Study Based on Sleep Apnea Detection, in Accepted in IEEE Congress on Evolutionary Computation2012: Brisbane, Australia.*
17. *Schluter, T. and S. Conrad, An approach for automatic sleep stage scoring and apnea-hypopnea detection. Frontiers of Computer Science, 2012. 6(2): p. 230-241.*
18. *Aksahin, M; Aydin, S.; Firat, H.; Erogul, O, Artificial Apnea Classification with Quantitative Sleep EEG Synchronization. Journal of Medical Systems, 2012. 36(1): p. 139-144.*
19. *Guijarro-Berdinas, B., E. Hernandez-Pereira, and D. Peteiro-Barral, A mixture of experts for classifying sleep apneas. Expert Systems with Applications, 2012. 39(8): p. 7084-7092.*
20. *Vapnik, V., Statistical Learning1998: Wiley.*
21. *Lauer, F. and G. Bloch, Incorporating prior knowledge in support vector machines for classification: A review. Neurocomputing, 2008. 71(7-9): p. 1578-1594.*
22. *Eberhart, R. and J. Kennedy. A new optimizer using particle swarm theory. in Micro Machine and Human Science, 1995. MHS '95., Proceedings of the Sixth International Symposium on. 1995.*
23. *Kennedy, J., R. Eberhart, and Ieee, Particle swarm optimization. 1995 Ieee International Conference on Neural Networks Proceedings, Vols 1-61995. 1942-1948.*
24. *Clerc, M. The swarm and the queen: towards a deterministic and adaptive particle swarm optimization. in Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on. 1999.*
25. *Engelbrecht, A.P., Fundamentals of Computational Swarm Intelligence2005: Wiley.*

Virtual Prosthodontic Planning for Oral Rehabilitation: a Pilot Study

Jaafar Abduo[1]

[1]School of Dentistry, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009
{jaafar.abduo@uwa.edu.au}

Abstract. This report illustrates the application of digital work-up as a dental treatment planning tool for prosthodontic patients. A clinical case was selected according to the prosthodontic needs. The conventional work-up was performed in accordance with the traditional prosthodontic criteria. The digital work-up was comprised of scanning of the pre-treatment models followed by restoring the deficient teeth and replacement of the missing teeth. The two diagnostic work-up models were compared in relation to the modified teeth, soft tissue and occlusal contacts. It appears that the digital work-up is a feasible option and allows quantitative and qualitative evaluation of the dental treatment impact.

Keywords: diagnostics, work-up, dental rehabilitation, dental prosthesis, image registration

1 Introduction

Many patients suffer from dentally related problems such as missing or deficient teeth, disorganized dental arches, or unacceptable tooth contour. Consequently, dental clinicians commonly encounter situations were alteration in tooth contour is indicated. Prosthodontics is the dental specialty that deals with the diagnosis, treatment planning, rehabilitation and maintenance of the oral function, comfort, appearance and health of these patients [1]. In many instances, prosthodontic treatment necessitates irreversible alterations to the remaining hard tissues. In order to justify such alterations, significant benefits of the treatment should be apparent. Therefore, to reach a satisfactory outcome, comprehensive diagnostic planning and work-up should be conducted before embarking on the definitive prosthodontic rehabilitation. Further, the diagnostic work-up allows visualizing the outcome and helps in deciding on the most adequate treatment plan for a specific case.

For dental abnormalities, the conventional prosthodontic protocol involves obtaining diagnostic models that represent the patient's dental arches upon which the diagnostic work-up can be performed. The complexity of the treatment ranges from single or few teeth restorations, to the complete dentition. The planned treatment can involve altering the tooth morphology, altering the vertical dimension of occlusion, reorganizing the occlusion and restoring all the teeth of at least one dental arch [2, 3]. In the dental laboratory, the diagnostic work-up involves preparing dental models, reducing part of the teeth and building the contours with wax [2, 4]. The Ideal diagnostic work-up should be applicable, transferable, conservative and aesthetic. Three critical criteria must be fulfilled. (1) The dental modifications should preserve the emergence profile. Therefore, although the dental modifications can be significant at the incisal or occlusal surfaces, they should be less prominent closer to the soft tissues. Further, excessive reduction should be avoided as this can affect the health of dental pulp. (2) The soft tissues should not be altered as they will aid in transferring the diagnostic information. (3) The occlusal contacts should be accurately located against the opposing teeth.

The outcome of this "trial" treatment can be demonstrated to the patient for approval or suggestion of any further modifications. In this manner, the patient will be more informed of the final outcome. Subsequently, the diagnostic work-up will facilitate the "outcome based treatment" which implies that the magnitude of irreversible alteration to the teeth is dictated by the final outcome rather than the initial patient presentation [5, 6]. This is accomplished clinically by preparing the teeth according to the anticipated final prostheses design as determined by the diagnostic work-up. In addition, provisional prostheses can be fabricated following the diagnostic work-up and, should the provisional outcome satisfy the patient, the definitive prostheses will be fabricated to resemble the diagnostic work-up [5, 6].

More recently, with the advent of laser scanning, virtual planning, rapid prototyping and computer-aided design and manufacturing, it is hypothesized that diagnostic planning can accomplished in time-efficient and well-controlled fashion. This pilot study introduces an additional application of computerized technologies that allows virtual alteration of dental morphology. It is envisioned that the digitally modified dentition can be used by the dental clinician and technician as a guide for the final prosthesis. In addition the study evaluates the potential of digital work-ups in producing acceptable outcome and compares the accuracy digital work-up to conventional work-up.

2 Materials and Methods

Human research ethics approval was obtained from the Human Research Ethics Committee of the University of Western Australia (RA/4/1/5097). A clinical case that requires diagnostic work-up prior to prosthetic treatment was selected. A total of 9 teeth in the maxillary arch required treatment. 6 teeth required restoration and 3 teeth required replacement. Maxillary and mandibular dental impressions were taken by irreversible hydrocolloid impression material (Alginate, GC America, IL, USA). The impressions were poured by dental stone (Buff Stone, Adelaide Moulding & Casting Supplies, South Australia, Australia). These models comprised the pre-treatment situation. Each model was duplicated twice by reversible hydrocolloid material (Megafeel, MKM System, Haanova, Slovakia). One set of models were treated by conventional work-up and the other by digital work-up.

2.1 Conventional Work-Up

The actual models were articulated on semi-adjustable articulator (Whip Mix, Louisville, KY, USA) and modified by trimming and contour alterations with inlay wax (VITA Zahnfabrik, Bad Sackingen, Germany). The required work-up modifications were even bilateral occlusal contacts, symmetry between the two sides, and natural teeth morphology (Fig. 1).

The conventional work-up models were scanned by Micro-CT scanner (SkyScan, Bruker, Kontich, Belgium). Virtual 3D Stereolithography (STL) images of the maxillary and mandibular models were constructed from the Digital Imaging and Communication Medicine (DICOM) images with the aid of DICOM viewing program (InVesalius, Renato, Archer Technology of Information Centre, Campinas, São Paulo, Brazil).

Fig. 1. Top left: Pre-treatment models. Top right: Conventional work-up models. Bottom: Scanned conventional work-up models.

2.2 Digital Work-Up

The maxillary and mandibular pre-treatment models were scanned by the Micro-CT scanner and STL images were constructed. A 3D rendering software package (Geomagic Studio, Raindrop Geomagic Inc., Research Triangle Park, NC, USA) was used to complete the digital work-up. The maxillary and mandibular models were virtually articulated by using the point-to-point alignment feature of Geomagic Studio. To obtain aesthetic tooth morphology, physiological teeth moulds (Phonares Teeth, Ivoclar Vivadent AG, Schaan, Liechtenstein) were scanned by the Micro-CT scanner. Each virtual tooth was fitted manually on the model with the aim of obtaining ideal teeth arrangement, emergence profile, symmetry and aesthetics. This is followed by ensuring ideal occlusal contacts exist. The virtual tooth alignment involved size alteration, rotation and translation (Fig. 2).

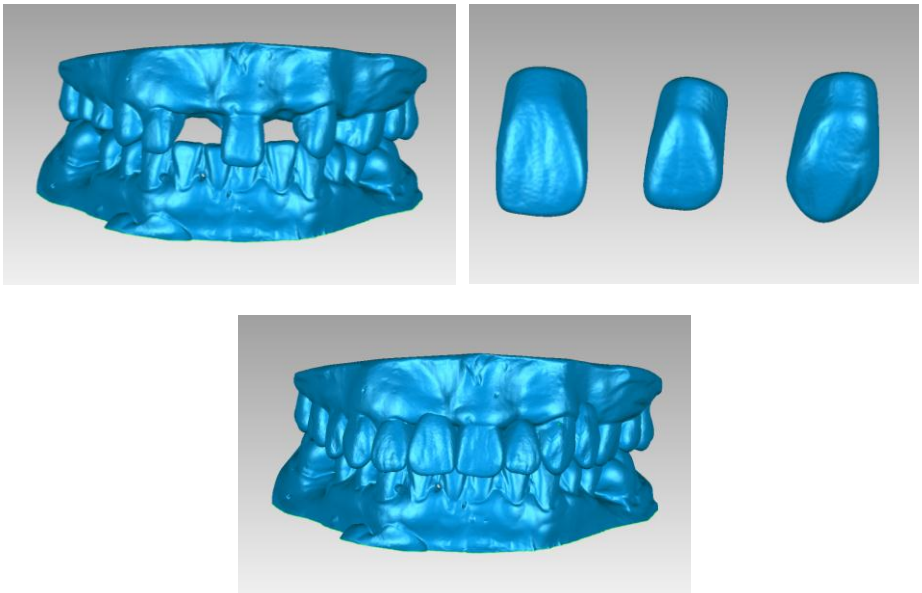


Fig. 2. Top left: Scanner pre-treatment models. Top right: Scanned physiological teeth. Bottom: Digital work-up models.

2.3 Analysis

The comparison between each treatment protocol was performed digitally at three levels; (1) tooth level, (2) soft tissue level, and (3) occlusion level. For the tooth level, alterations in the contour are expected as long as the emerging portion of the tooth is minimally affected and excessive reduction is avoided. Since the soft tissues are used as landmarks, they should not be altered. The occlusal contacts should be well-distributed and of comparable magnitude.

To compare the tooth morphology, each work-up models was superimposed on the pre-treatment model by the process of image registration. The image registration is comprised of point-to-point registration followed global registration. Eventually, the models were aligned through the best fit principles according to Iterative Closest Point (ICP) algorithm [7, 8]. This step aimed to evaluate the amount of tooth modifications that will be applied by each diagnostic work-up. The discrepancy distribution between the superimposed models was illustrated in colour-difference maps to locate the dimensional positive and negative deviations. The threshold value was set at 2 mm. The warm colours represent positive deviations, while the cold colours represent negative deviations. The green colour indicates an optimal match.

The soft tissue band closer to the teeth was trimmed from each model and superimposed on the pre-treatment model. In addition to colour-difference maps generation, the average 3D Euclidean Distance (ED) of 2000 random points of the common surfaces of the two models was calculated. The absolute deviation values were used to solely quantify the deviation magnitude. Therefore, the less mean distance between the models, the better the accuracy of the diagnostic work-up.

The occlusal contacts were measured by Meshlab Software (Visual Computing Lab, University of Pisa, Italy). This was applied by measuring the distance between the occlusal surface of each maxillary model and mandibular model. The threshold value was set at 1 mm. According to the distance, the occlusal surface of each maxillary model was colour coded. The red colour indicates contact and the blue colour indicates lack of contact. Table 1 illustrates the comparison criteria of each level.

3 Results

3.1 Teeth Comparison

Following the image registration of each work-up model, the unrestored teeth were closely matched. This confirmed the accuracy of the registration process.

In relation to the conventional work-up (Fig. 3), the occlusal plane and teeth alignment were even. The symmetry between each side was observed. In general, the restored teeth were enlarged. Emergence profile was preserved and the teeth gradually enlarged toward the occlusal and incisal surfaces, where the enlargement was most excessive. The central incisor had more reduction (about 1 mm) on its facial aspect. This amount of reduction appears to be within the clinical acceptability.

The digital work-up (Fig. 4), had similar presentation in relation to occlusal plane, teeth alignment and symmetry. In general, the modified teeth exhibited greater definition than conventional work-up which can contribute to more aesthetic appearance. The emergence profile and enlargement features are similar to the conventional work-up. The amount of reduction on the facial aspect of the central incisor was less prominent than for the conventional work-up.

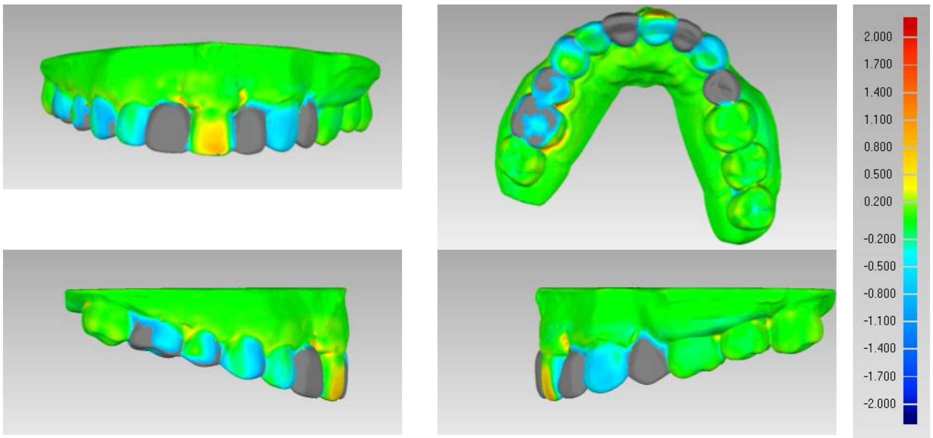


Fig. 3. Colour-difference maps of the conventional work-up model after superimposition on pre-treatment model.

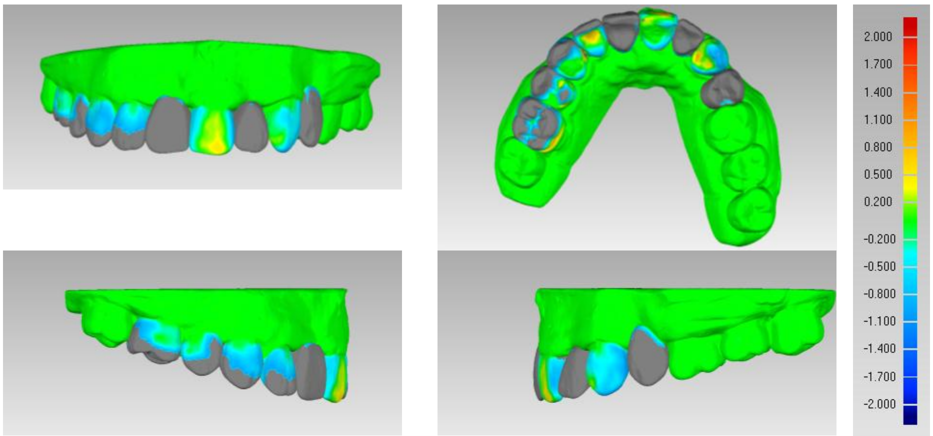


Fig. 4. Colour-difference maps of the digital work-up model after super-imposition on pre-treatment model.

3.2 Soft Tissue Comparison

The average ED between the pre-treatment and the conventional work-up models was 0.152 mm (SD = 0.06 mm). Greater average ED value was observed between the pre-treatment and the digital work-up models (0.779 mm, SD = 0.856 mm). This indicates that alteration of the soft tissue contour by the conventional work-up is less than the alteration by the digital work-up. The colour-difference maps illustrated that greater discrepancy magnitudes exist closer to the teeth for digital work-up model (Fig. 5).

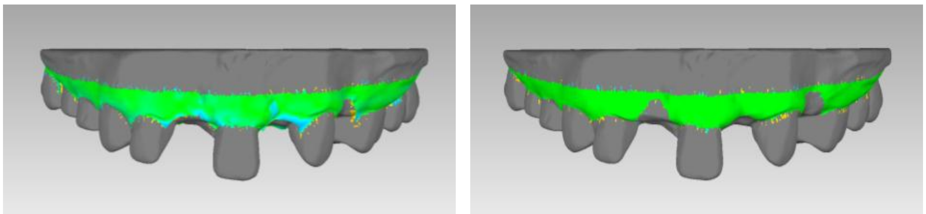


Fig. 5. Colour-difference maps of segmented soft tissue band following superimposition on pre-treatment model. Left: Conventional work-up model. Right: digital work-up model.

3.3 Occlusion Comparison

For the two work-up models, all the opposed teeth were in contact with the mandibular teeth (Fig. 6). It appears that the contact for the conventional work-up was more even in magnitude and distribution. Still, both of them improved the occlusal contacts in comparison with the pre-treatment model.

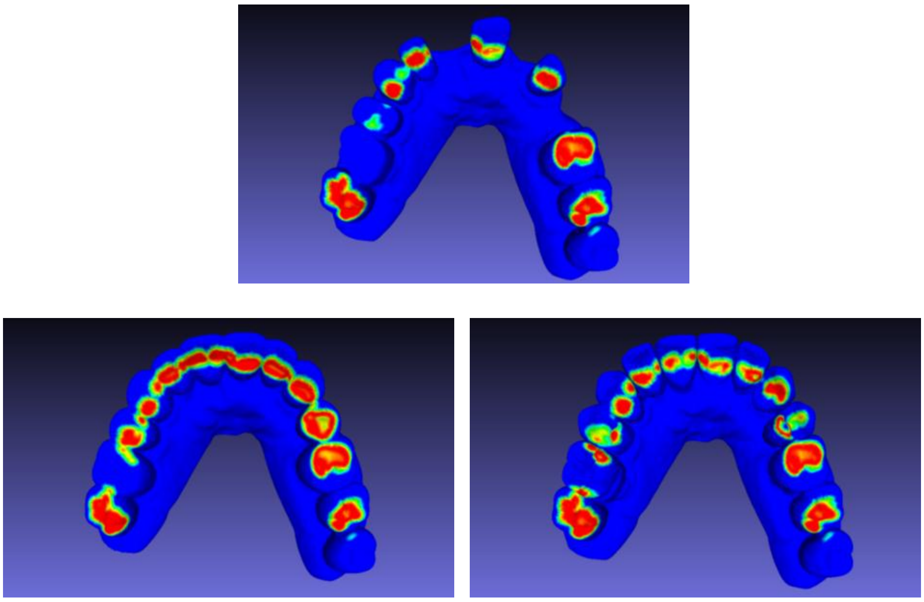


Fig. 6. Occlusal contacts on the maxillary teeth. Top: Pre-treatment model. Bottom left: Conventional work-up model. Bottom left: Digital work-up model.

4 Discussion

Today, with patients' high aesthetic expectations, the possible high standard of treatment and the risk of legal litigations, more emphasis has to be placed on the diagnostic work-up to allow the patient to visualize the final outcome. On the basis of accurate planning, the patient will be able to provide consent after being fully informed by observing the diagnostic work-up outcome. However, it is common for the clinician to omit the diagnostic work-up procedure, or to leave it to dental technicians to decide. The reason behind this is that the conventional diagnostic work-up is a time consuming process and requires special training and artistic abilities. The technicians might have the expertise in developing natural looking dental morphology; however, they commonly lack the visualization of the biological parameters. Therefore, the predictable application of the diagnostic work-up is limited to clinicians with a special level of training.

It is speculated that the introduced digital work-up approach will alleviate many of the difficulties associated with conventional work-up. In addition, digital work-up exhibits several advantages that justify its routine application. For example, the process is completed virtually and requires no physical material which has significant economical implication. The pre-treatment models are not altered and can be preserved in the patient record. The applicability and efficiency of this technique are further augmented by automation of the tooth modification process. The evaluation of the implications of the diagnostic work-up is a useful feature to analyze the feasibility of the proposed treatment [9]. It is expected that this feature will enhance patient's communication and reviewing dental modifications with relative ease.

In the era of digital dentistry, it is more likely in the future that digital dental impression will become popular [10]. Therefore, models manipulation will be purely virtual, which will omit the conventional impression taking, model pouring and subsequent scanning. To enhance the applicability of the digital work-up, actual models can be produced by rapid prototyping technology. The printed model is envisioned to provide direct guidance to the involved dental clinician and technician.

The digital protocol appears to be promising and very comparable to conventional protocol. In fact the high tooth definition implies that the computerized protocol can produce more aesthetic treatment planning. However, the tooth-soft tissue junction appears to be more affected with the digital work-up than conventional work-up. This was illustrated by the greater ED for the digital work-up. Certainly, greater soft tissue accuracy is always desirable, although the clinical significance of the reported discrepancies cannot be confirmed.

5 Conclusion

This pilot study confirms the applicability of digital treatment planning and tooth modifications for patients presenting with deficient or missing teeth. In addition, the proposed approach allows the quantitative and qualitative diagnostic evaluations of the impact of dental treatment on the existing dentition. In relation to accuracy, the conventional protocol is more accurate at soft tissue and occlusion levels; however this observation should be validated with greater sample number.

Acknowledgements. The author thanks Professor Mohammed Bannamoun for his guidance on the project. The author acknowledges the facilities, the scientific and technical assistance of the National Imaging Facility at the Centre for Microscopy, Characterization & Analysis, University of Western Australia, a facility funded by the University, State and Commonwealth Governments. This paper is part of a project supported the Research Development Award from the University of Western Australia and the Raine Priming Grant.

References

1. The Glossary of Prosthodontic Terms. J. Prosthet. Dent. 94, 10-92 (2005)
2. Keough, B.: Occlusion-Based Treatment Planning for Complex Dental Restorations: Part 1. Int. J. Periodontics. Restorative. Dent. 23, 237-247 (2003)
3. Besimo, C.E., Rohner H.P.: Three-Dimensional Treatment Planning for Prosthetic Rehabilitation. Int. J. Periodontics. Restorative. Dent. 25, 81-87 (2005)
4. Abduo, J.: An Innovative Prostheses Design for Rehabilitation of Severely Mutilated Dentition: a Case Report. J Adv. Prosthodont. 3, 37-42 (2011)
5. Magne, P., Belser, U.C.: Novel Porcelain Laminate Preparation Approach Driven by a Diagnostic Mock-Up. J. Esthet. Restor. Dent. 16, 7-16 (2004)
6. Gurel, G.: Porcelain Laminate Veneers: Minimal Tooth Preparation by Design. Dent. Clin. North. Am. 51, 419-431 (2007)
7. Williams, J., Bennamoun, M.: Simultaneous Registration of Multiple Corresponding Point Sets. Comput. Vis. Image. Und. 81, 117-142 (2001)
8. Mian, A.S., Bennamoun, M., Owens, R.: Three-Dimensional Model-Based Object Recognition and Segmentation in Cluttered Scenes. IEEE. T. Pattern. Anal. 28, 1584-1601 (2006)
9. Davis, G.R., Tayeb, R.A., Seymour, K.G., Cherukara, G.P.: Quantification of Residual Dentine Thickness Following Crown Preparation. J. Dent. 40, 571-576 (2012)
10. van Noort, R.: The Future of Dental Devices is Digital. Dent. Mater. 28, 3-12 (2012)

Development of Patient-Practitioner Assistive Communications (PPAC) Ontology for Type 2 Diabetes Management

David Forbes[1*], Pornpit Wongthongtham[1], Jaipal Singh[2]

[1]School of Information Systems
[2]Department of Electrical and Computer Engineering
Curtin University, Perth, Australia
david.e.forbes@postgrad.curtin.edu.au,
P.Wongthongtham@cbs.curtin.edu.au, j.singh@curtin.edu.au

Abstract. Communication in primary care is a key area of healthcare slow to adopt new technology to improve understanding between the patient and healthcare practitioner. Patients whose cultural background and regular form of dialectal communication are far removed from that of mainstream society are particularly disadvantaged by this during the patient-practitioner interview encounter (PPIE). In this paper, we present an assistive communications technology (ACT) framework for PPIE developed using a Type-2 Diabetes Management Patient-Practitioner Assistive Communications (T2DMPPAC) ontology in order to help both Aboriginal patient and non-Aboriginal practitioner optimise their pre-encounter, during-encounter and post-encounter communication. The T2DMPPAC architecture provides knowledge and presents it in a manner that is easily accessible and understood by the user (patients and practitioners) as well as accompanying carers, and as appropriate, interpreters. An example of bi-directional mapping of concepts to language during a PPIE session is shown using the ontology.

Keywords: Type-2 diabetes management, ontology, assistive communication, Aboriginal English pragmatics

1 Introduction

Communication in primary care settings is one of the key areas of healthcare that is often slow to adopt new technology to improve understanding between the patient and healthcare practitioner. Previous work has shown extreme weakness in communications between practitioners and patients, particularly for patients whose cultural background and regular form of dialectal communication is far removed from that of mainstream society [1-3]. The use of the description 'practitioner' covers the medical professional who works directly with patients as a provider of healthcare. The focus of this paper is on primary care and specifically on face-to-face patient-practitioner interview encounters, which for convenience we term 'PPIE'.

The differences in cognitive capabilities, age factored illness, and cultural communication disparities together with PPIE time constraints, place a very high expectation of expertise and effectiveness on the practitioner when interacting with a patient. Rightly so as the practitioner, being the expert in his/her field, will have to determine the course of action to take to ensure the wellbeing of the patient.

However, the enormity of communicating the relevant information, of variable interpretations in conversation and in unpredictable contextual circumstances, brings its own complexity and risk of misunderstanding. Furthermore, tracking the history of the patient and assigning reliable meaning to occurrences should not just be limited to interactions in the PPIE. In this research, attention is paid to the potential value of pre-encounter and post-encounter communications surrounding the PPIE so that the patient is empowered with knowledge and prepared to contribute towards his/her own healthcare.

The calls for more education and training of providers in human skills as well as healthcare knowledge fails to acknowledge that expectations of advancement cannot be realised without technological tools to aid in primary healthcare. Health care is an information and knowledge intensive industry; but ICT investments found elsewhere are virtually absent in the primary care communications protocols.

In 2007, Kaiser Permanente's Southern California Region introduced a program named the Proactive Office Encounter (POE), to address the growing large scale patient need for preventive care and management of chronic disease[4]. However, the POE imposes a fairly high degree of labour intensity in spite of the inclusion of electronic information systems. The pre-encounter and post-encounter functions are carried out by nursing staff. The authors of a paper on the POE [4] make it clear that for optimal benefit, the POE will require modified processes, structures and management work changes employing smart tools. Our research brings this thought process into play with the added complication of intercultural health care communication. To this end, we have identified computer ontologies as providing the most versatile means to equip a form of assistive communications technology (ACT) to help both patient and practitioner communicate better so as to improve diagnosis and compliance for more beneficial healthcare outcomes.

The need for practitioners to communicate more effectively, particularly to patients from different cultural and language backgrounds is taking on more importance as the intake of refugee migrants from non-English speaking backgrounds increases. The English language proficiency of members within the minority groups in Australia are varied, with some being very proficient in English while others being quite poor. Thus, confronted with patients who are disadvantaged through the cultural disconnect of significant differences in ethnic values and practices; western dominated health literacy; and language/dialect, practitioners are also disadvantaged, by the lack of effective support systems that can counter these handicaps.

In terms of inter-cultural communications in healthcare, the Aboriginal history in and experiences of westernised health care interaction barriers has provided Australia with a strong but under-utilised grounding in the challenges that health care providers face when negotiating the health service needs of ethnic minorities. We assert that

Aboriginal English Home Talk (AEHT) can be used as a model for ethnic minority immigrant communications acculturation.

In this paper, we focus on the chronic disease type 2 diabetes mellitus in Aboriginal people. The evidence across the diverse cultures suggest that the Aboriginal community and other minorities are willing to trust the treatments they receive if the practitioners explains to them why and how a particular test, course of treatment and care plan is the best way to improve their quality of life; life expectancy; and self-management of the chronic disease condition type 2 diabetes mellitus (T2DM) or any other chronic disease for that matter [5-8]. This is an extremely demanding scenario and poorly managed Type 2 Diabetes Mellitus (T2DM) consequences are too often witnessed in emergency department and hospital admissions due to the failure of providing proper explanation to the patient.

The first step is to understand the implications of chronic disease T2DM. It is an incurable disease condition that can be managed. Such patient groups often fall short on adherence to medical advice and we attribute this to poor communication due to a combination of cognitive and health literacy barriers [9-12]. We consider that while practitioner training, re-training, cultural education and other supportive measures that include interpreter services are worthy, they are of very limited effect when the scale of the communications complexities and growth of chronic disease patients are factored in.

The use of assistive technology in particular of development of Type 2 Diabetes Management Patient-Practitioner Assistive Communications (T2DMPPAC) ontology is intended to augment these elements and optimise opportunity for patients, practitioners, carers and interpreters to share in a community knowledge capture and health literacy development set of tools. As explained by Gruber, 'ontology defines a set of representational primitives with which to model a domain of knowledge or discourse'[13]. These are typically classes and their attributes / properties, and describe/qualify relationships among class members. Definitions include information such as annotations about meaning; and constraints on logical consistency in application. Domain ontologies can be mapped to other domain ontologies, thereby presenting the opportunity to create greater interactivity and versatility involving hitherto underdeveloped or non-existent discourse concepts and schemas.

The collision of clinical language and established western medical practice, with long established non-clinical and non-westernised cultures and modes of communications is a serious challenge to effective engagement pragmatics and health outcomes. The advantage presented to us through ontology development is accentuated by the existence and continued advancement of the Semantic Web, opening up the possibilities for independent and individual access, sharing and reuse of ontology supported communications systems via the increasingly ubiquitous smart devices such as mobile telephones and portable touchscreen tablets. The latter directly proffer pre-encounter and post-encounter, plus PPIE input, versatility.

This paper is comprised of sections. A review of previous work in developing ontologies for type-2 diabetes is presented in section 2. The Patient-Practitioner Assistive Communications architecture is shown in section 3 while the structure and usage of the Type-2 Diabetes Management Patient-Practitioner Assistive Communications is shown in section 4. This paper concludes in section 5.

2 Literature Review

The term 'Ontology' is derived from its usage in philosophy where it means the study of being or existence as well as the basic categories [14]. Therefore it is used to refer to what exists in a system model. In computer science, ontology is the effort to formulate an exhaustive and rigorous conceptual schema within a given domain, typically a hierarchical data structure containing all the relevant concepts and relationships between those concepts. In artificial intelligence, ontology is an explicit specification of a conceptualisation [15, 16].

In this research an ontology is a domain knowledge representation formed upon a controlled, standardised vocabulary for describing classes and the semantic relationships between them. The T2DPPAC ontology aims to overcome communication barriers due to culture gaps between practitioner and Aboriginal patient. Hence in the ontology, a standardised vocabulary drawn from type-2 diabetes management guidelines is captured along with Aboriginal English home talk. The Aboriginal diabetic patient uses the ontology to understand diabetic concepts in their Aboriginal discourse. The practitioner and involved people e.g. interpreter, use the ontology to understand Aboriginal culture and find a way to communicate with the patient.

There are researchers putting effort towards diabetes ontology development. Chalortham et al. developed diabetes mellitus ontology which covers risk assessment, diagnosis and complication, treatment, and follow-up [17]. Based on the ontology reminding system was developed as part of type 2 diabetes mellitus clinical support system. The diabetes mellitus ontology was developed based on Thailand Diabetes Mellitus Clinical Practice Guideline 2008 and suggestion of medical experts. Buranarach et al. introduced the synopsis of chronic disease healthcare framework in which the important of ontology for healthcare knowledge management system was pointed out [18]. Lin and Sakamoto developed Glucose Metabolism Disorder ontology which was classified into diabetes mellitus, diabetes complication, hyperglycaemia, hyperrinsulinism, etc. [19]. The ontology was also linked to Geographical regions ontology and Genetic Susceptibility Factor ontology to describe the genetic susceptibility factors to Diabetes Mellitus. Ganendran et al. developed ontology based multi-agent systems in which diabetes management was applied as a case study involving three agents i.e. specialist agent, patient agent, and web agent [20]. Shahar et al. developed Knowledge Based Temporal Abstraction (KBTA) focusing on shared knowledge representation and reuse [21]. However, none of work focuses on assistive communications particularly for ethnic minority immigrant communications acculturation. In addition there is no existing T2DM ontology developed based on Australian recognised professional healthcare standard guidelines.

There are a number of ontology methodologies including NeOn, Knowledge Engineering, DOGMA, TOVE, Methontology, SENSUS, DILIGENT, etc. NeOn methodology is a scenario based methodology that provides direction for all key aspects of the ontology engineering process [22]. In contrast to other methodologies those provide methodological guidance for ontology engineering, the NeOn methodology does not suggest a rigid workflow but it prescribes pathways instead as well as processes and activities for a variety of scenarios [23]. The nine scenarios identified in the NeOn methodology are for ontology engineering and special emphasis is placed on reusing and re-engineering knowledge resources both ontological and non-ontological [24]. We use the NeOn methodology as it provides the most flexibility for development of the ontology. The tool used in the implementation process is protégé 4.2.

3 Patient-Practitioner Assistive Communications (PPAC) Architecture

This section provides a new approach to using ICT technology for enabling patient-practitioner assistive communications (PPAC) in primary care. The PPAC architecture shown in Figure 1 is an ontology based system that provides knowledge and presents it in a manner that is easily accessible and understood by the user (patients and practitioners). In this paper, we use type-2 diabetes as the health domain that is represented by a type-2 diabetes management (T2DM) ontology. The T2DM ontology is developed from scratch using available non-ontological resources namely the Royal Australian College of General Practitioners (RACGP) T2DM Guidelines for management of Type 2 diabetes. An ontology to represent the Aboriginal language, in this case Aboriginal English Home Talk, is also developed from scratch. Aboriginal words and phrases which are gradually populating the ontology include contributions from the members of an Aboriginal Nyungar focus group of trainee nurses who gathered to assist the authors in April of 2011. The work of this focus group was led by a moderator who used the RACGP T2DM Guidelines to help validate mappings of semantics between the clinical English diagnosis processes and the undocumented pragmatic cultural expressions. In other words, the guidelines provided an orderly track to prompt discussion that sought responses that included Aboriginal English words, phrases and advisory explanations.

Australia's Aboriginal cultural history exceeds 40,000 years yet little is documented in health care literature that accommodates the unique characteristics of beliefs, perceptions and practices to help deliver effective care and wellbeing outcomes. A substantial source of Aboriginal English Home Talk research literature, emanating from the field of education as opposed to medicine or healthcare is being employed to bolster the communications gaps through this ontology development. This work is unique as there has not been a representation of Aboriginal English for medicine in an ontology. The ontology will provide a mapping of terms from the formal clinical T2DM to Aboriginal English, allowing improved two-way

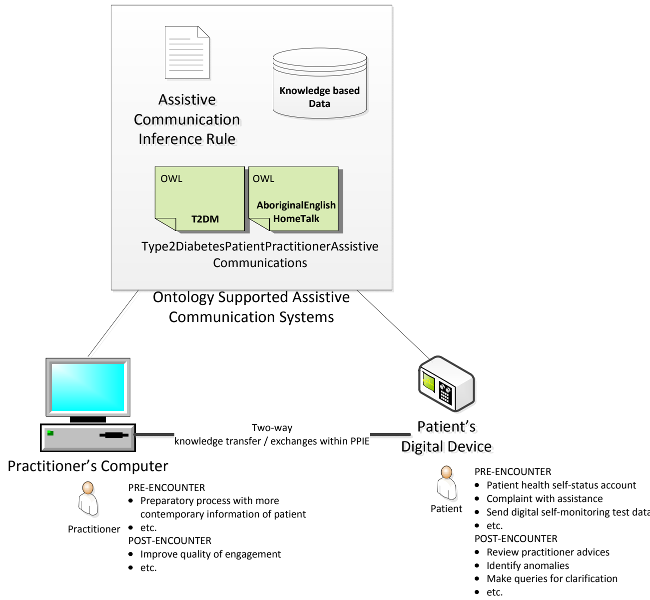


Fig. 1. Patient-Practitioner Assistive Communication Architecture for Type-2 Diabetes Management

communication between the patient and practitioner through the ontology intermediary. The T2DMPPAC ontology, comprised of the T2DM and AEHT ontology vocabularies, effectively surrounds the PPIE and empowers the patient but does not intrude on clinical skills. Its use will help to overcome the barriers in communication by facilitating patient PPIE preparation and post-PPIE review, which implicitly brings other shared knowledge benefits. This modular architecture allows for different language ontologies or health care ontologies to be used in any healthcare setting, thus providing a robust tool for improving communication and understanding during the PPIE for other medical conditions.

3.1 Stages in Patient-Practitioner Interview Encounter

While previous work was concerned with the PPIE in the context of interaction, we found that breaking this process into three phases, pre-encounter, during-encounter and post-encounter, actually aids in better communication and understanding for both patient and practitioner.

The pre-encounter enhances PPIE effectiveness through separate preparations by the patient and by the practitioner. Preparation for the patient is a mental exercise in rehearsing the intended patient health self-status account or complaint through a digital assistant that may also become an educational process and an opportunity to send digital self-monitoring test data. Data enrichment can then elevate the practitioner's preparatory process with more contemporary information than merely the record of prior consultation. The difference is one of timing, in that practitioner preparation may be minutes before the PPIE and the patient's preparation is likely to be many hours or days before.

In the during-encounter stage, the practitioner will collect information from the patient, clarify ambiguity in information, examine the patient, decide on action to be taken and present course of action to the patient. The information collected during the pre-encounter stage will greatly improve communication between patient and practitioner as both sides have a better understanding of the ailment. However, the final decision rests with the practitioner on the nature of the ailment and action required to address it.

Acknowledging the fact that detail of the exchanges within the PPIE are not so easily or reliably recalled by the patient as opposed to almost instant data entry by the practitioner(s), a post-encounter facility to accommodate a patient review can add to that educational value. It can also facilitate improved quality of further practitioner engagement by allowing the patient to identify anomalies, raise queries for clarification and even influence the mode of engagement by the practitioner in future PPIEs.

The ontology aids all three processes by providing a comprehensive set of standards and guidelines for patients and practitioners to follow as well as present it in a manner that is comprehensible to all users.

4 Type-2 Diabetes Management Patient Practitioner Assistive Communications (T2DMPPAC) Ontology Development

This section aims to show how we built the T2DMPPAC ontology. There are two main parts in the ontology i.e. Type 2 Diabetes concepts which classify all concepts related to type 2 diabetes and Aboriginal English Home Talk concepts which classify all concepts used in Aboriginal communications as shown in Figure 2. These two main parts are formed into two main ontology classes which are linked together through ontology relations and constraints. The two classes are self-standing concepts which we form them as sub classes of class Independent_Concept.

The relations i.e. object properties mapped the two classes are inAboriginalEnglishHomeTalk and inType2DiabetesConcept in which they are inverse to each other. Figure 3 shows relations between classes Aboriginal_English_Home_Talk and Signs_and_Symptoms through object properties inAboriginalEnglishHomeTalk and inType2DiabetesConcept.

4.1 Case Study: Aboriginal Patient With Blurred Vision

We provide a simple case study to illustrate how the ontology can be used in PPIE. Diabetes patients with high blood sugar may suffer from blurred vision. This might be a temporary condition or a precursor to more serious conditions such as retinopathy, glaucoma or cataract. An Aboriginal patient may walk into a clinic once he/she notices their vision is blurred. Typically, the patient will say they have "bad eyes" when seeing the doctor or they might choose to say “Gooras Winyarn”, which is Aboriginal English for blurred vision. If the condition is serious, the patient would use the Aboriginal English word as it not only provides a description of the problem but also the severity of it, which is not captured in Standard Australian English.

Knowledge captured in the T2DMPPAC ontology as shown in Figure 3 illustrates that the Nyungar Aboriginal words of gooras winyarn can be taken to mean blurred vision or altered vision. Literal translation between traditional or original Aboriginal words that now have a place within Aboriginal English pragmatics is limited and not always sufficiently explicit as to carry a specific meaning. Such words appear in phrases and accord with circumstance; and will therefore vary in context. It is not appropriate for instance, to assign a distinct Australian English oriented meaning to ‘gooras winyarn’ unless the context is completely clear. In the situation where a patient is anxious and/or has limited English proficiency, such words become key triggers to justify ontological system queries. The practitioner will know from the annotation that together these words come close to meaning ‘eyes bad’, thereby informing the practitioner that an eye problem is suspected and will require investigation. The investigation may then include comparison with a prior PPIE record in addition to physical examination that will then better determine whether the condition is altered or blurred vision, or possibly both.

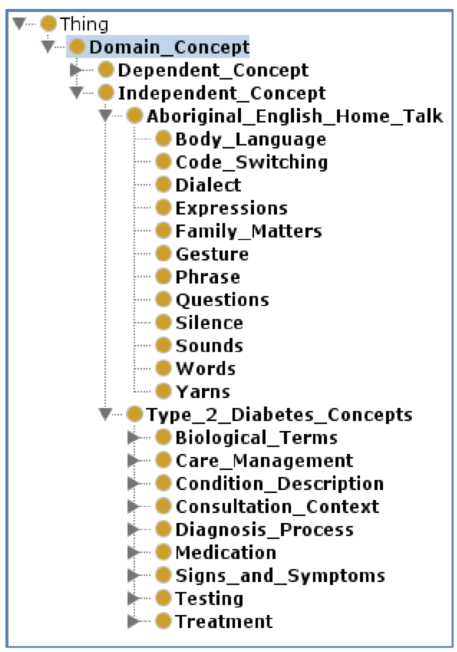


Fig. 2. Class Independent_Concept and its sub classes.

As can be seen in Figure 3, gooras_winyarn is an instance of class Words which is sub class of class Aboriginal_English_Home_Talk. Instance Altered_Vision which is same instance as Blurred_Vision is instance of class Vision i.e. sub class of class Signs_and_Symptoms i.e. sub class of class Type_2_Diabetes_Concepts. The instances are mapped through object properties inAboriginalEnglishHomeTalk and inType2DiabetesConcept.

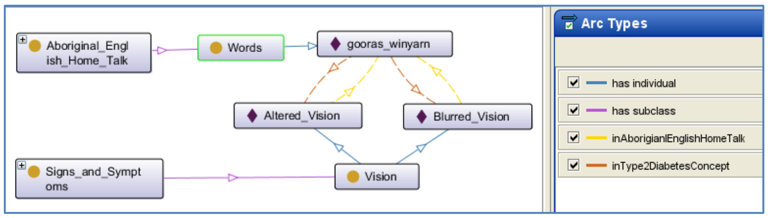


Fig. 3. Object properties inAboriginalEnglishHomeTalk and inType2DiabetesConcept mapped classes Aboriginal_English_Home_Talk and Signs_and_Symptoms

4.2 Refining the T2DMPPAC Ontology

There are refining concepts which will add meaning to other concepts. We form these into class Dependent_Concept as shown in Figure 4.

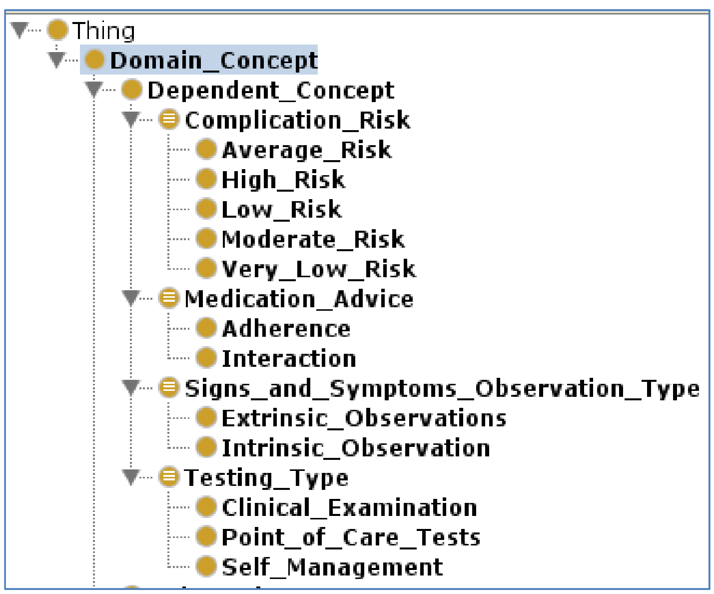


Fig. 4. Class Dependent_Concept and its sub classes

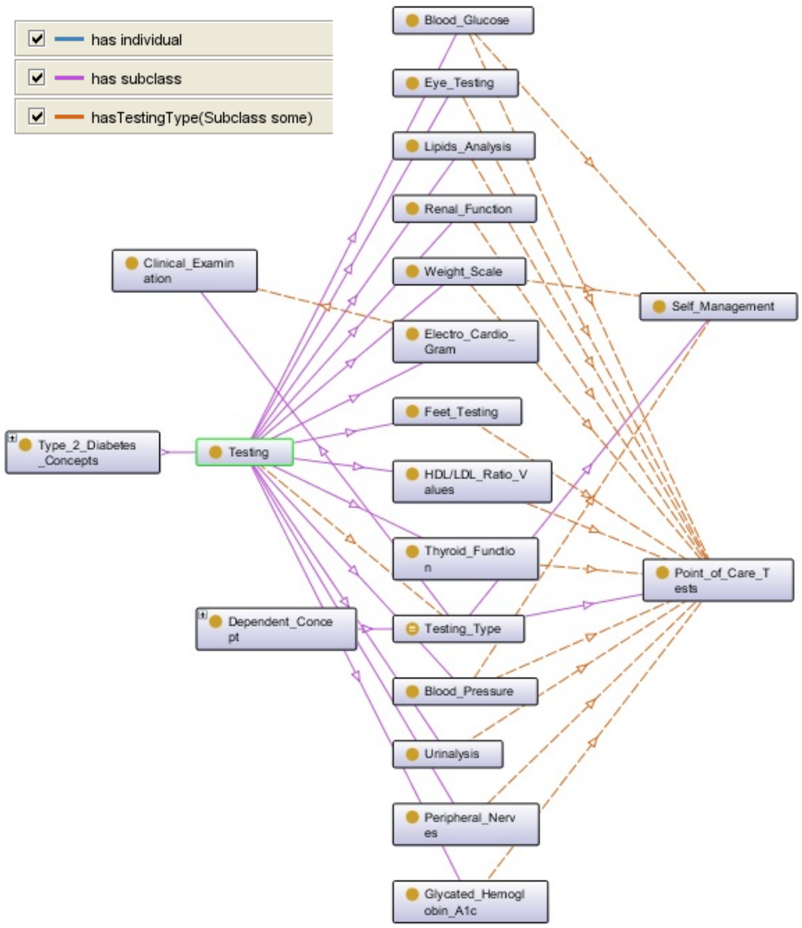


Fig. 5. Class Testing relates to class Testing_Type through object property hasTestingType

Class Complication_Risk adds risk value to any sub class of class Type_2_Diabetes_Concepts through relation hasComplicationRisk. It can be restricted to particular risk of average risk, high risk, low risk, moderate risk, or very low risk. Class Medication_Advice adds value in term of medication advice of adherence and/or interaction to any classes under classes Care_Management, Treatment, and Medication through relation hasMedicationAdvice. Observation of particular sign and symptom of patient can be specific to extrinsic or intrinsic observation of the patient. This can be specified through relation hasObservationType. Figure 5 shows class Testing_Type adding value to class

Testing through relation hasTestingType in term of types of testing i.e. clinical examination, point of care tests and self-management.

5 Conclusion

This paper introduces a novel approach to using ICT in the patient-practitioner interview encounter (PPIE). We developed a framework that links medical information with different language and cultural information to provide ease of understanding and communication between the patient from a minority group with a healthcare practitioner from a different cultural group. The key component of this framework is the Type-2 Diabetes Management Patient-Practitioner Assistive Communication Ontology. We showed how this ontology was created and the links between different classes and components in the ontology. We also presented a case study on how this ontology can be used by the Aboriginal patient to the practitioner.

For future work, we intend to populate the Aboriginal English ontology with as many medical words and phrases used within the Aboriginal community. We will then validate this ontology by the results it provides in a selection of typical PPIE situations faced by Aboriginal patients consulting non-Aboriginal practitioners.

6 References

1. Fortin, A.H.: Communication skills to improve patient satisfaction and quality of care. Ethnicity & disease 12, S3-58-61 (2002)
2. Johnson, R.L., Roter, D., Powe, N.R., Cooper, L.A.: Patient Race/Ethnicity and Quality of Patient-Physician Communication During Medical Visits. American Journal of Public Health 94, 2084-2090 (2004)
3. Kagawa-Singer, M., Kassim-Lakha, S.: A strategy to reduce cross-cultural miscommunication and increase the likelihood of improving health outcomes. Acad Med. 78(6) 78, 577-587 (2003)
4. Kanter, M., O. Martinez, G. Lindsay, K. Andrews, Denver., C.: Proactive Office Encounter: A Systematic Approach to Preventive and Chronic Care at Every Patient Encounter. The Permanente Journal 14, (2010 https://www.thepermanentejournal.org/files/Fall2010PDFS/ChronicCare.pdf)
5. Anderson, K., Devitt, J., Cunningham, J., Preece, C., Cass, A.: "All they said was my kidneys were dead": Indigenous Australian patients' understanding of their chronic kidney disease. MJA 189, 499-503 (2008)
6. Cass, A., Lowell, A., Christie, M., Snelling, P., Flack, M., Marrnganyin, B., Brown, I.: Sharing the true stories: improving communication between Aboriginal patients and healthcare workers. Medical Journal of Australia 176, 466-470. (2002)
7. Garvey, D.: A review of the social and emotional wellbeing of Indigenous Australian peoples – considerations, challenges and opportunities Edith Cowan University, Western Australia (2008. http://www.healthinfonet.ecu.edu.au/sewb_review Accessed 25 April 2012)
8. Kelly, J., Pekarsky, B., Dwyer, J., Mackean, T., Willis, E., Glover, J., Battersby, M.: Managing Two Worlds Together: Study 4 — Complex Country Aboriginal Patient Journeys., The Lowitja Institute (2011)

9. Benson, J.: Concordance--An alternative term to 'compliance' in the Aboriginal population. Aust Fam Physician 34, 831-835 (2005)
10. Hays, R.: Practising Rural Medicine in Australia. Eruditions Publishing, Melbourne, Australia (2002)
11. Lowell, A.: Communication and Cultural Knowledge in Aboriginal Health Care. A review of two subprograms of the Cooperative Research Centre for Aboriginal and Tropical Health's Indigenous Health and Education Research program. The Cooperative Research Centre for Aboriginal and Tropical Health, Casuarina NT (1998)
12. Schoen, D., Balchin, D., Thompson, S.: Health promotion resources for Aboriginal people: lessons learned from consultation and evaluation of diabetes foot care resources. Health Promotion Journal of Australia 21, 64-69 (2010)
13. Gruber, T.: Ontology. In: Liu, L., Özsu, M.T. (eds.) The Encyclopedia of Database Systems. Springer-Verlag (2009)
14. Witmer, G.: Dictionary of Philosophy of Mind - Ontology. Accessed on: 15 July 2010, from: http://philosophy.uwaterloo.ca/MindDict/ontology.html (2004)
15. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies 43, 907-928 (1995)
16. Gruber, T.R.: A translation approach to portable ontology specifications. Knowledge Acquisition 5, 199-220 (1993)
17. Chalortham, N., Buranarach, M., Supnithi, T.: Ontology Development for Type II Diabetes Mellitus Clinical Support System. In: Conference Ontology Development for Type II Diabetes Mellitus Clinical Support System. (Year)
18. Buranarach, M., Supnithi, T., Chalotham, N., Khunthong, V., Varasai, P., Kawtrakul, A.: A Semantic Web Framework to Support Knowledge Management in Chronic Disease Healthcare. In: Conference A Semantic Web Framework to Support Knowledge Management in Chronic Disease Healthcare. (Year)
19. Lin, Y., Sakamoto, N.: Ontology Driven Modeling for the Knowledge of Genetic Susceptibility to Disease. Kobe Journal of Medical Sciences 54, E290-E303 (2008)
20. Ganendran, G., Tran, Q.-N., Ganguly, P., Ray, P., Low, G.: An Ontology-driven Multi-agent approach for Healthcare. In: Conference An Ontology-driven Multi-agent approach for Healthcare. (Year)
21. Shahar, Y., Das, A.K., Tu, S.W., Kraemer, F.B., Musen, M.A.: Knowledge-Based Temporal Abstraction for Diabetic Monitoring. In: Conference Knowledge-Based Temporal Abstraction for Diabetic Monitoring, pp. 697-701. (Year)
22. Suárez-Figueroa, M.C.: NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse. Computer Faculty, vol. PhD. Universidad Politécnica de Madrid, Madrid, Spain (2010)
23. Villazón-Terrazas , B., Ramírez, J., Suárez-Figueroa, M.C., Gómez-Pérez, A.: A network of ontology networks for building e-employment advanced systems. Expert Systems with Applications 38, 13612-13624 (2011)
24. Villazón-Terrazas, B.: A method for reusing and re-engineering nonontological resources for building ontologies. Computer Faculty, vol. Ph.D. thesis. Universidad Politécnica de Madrid, Madrid, Spain (2011)

Three Dimensional Imaging Based Diagnosis for Obstructive Sleep Apnoea: A Conceptual Framework

[1]Syed M. S. Islam, [1]Mithran S. Goonewardene and [2]Paul Sillifant

[1] School of Dentistry, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia
[2] Department of Oral and Maxillofacial Surgery, Royal Perth Hospital, 197 Wellington Street, Perth, WA 6000, Australia
[1]{syed.islam,mithran.goonewardene}@uwa.edu.au, [2]paulsillifant@hotmail.com

Abstract. Obstructive Sleep Apnoea (OSA) is a disorder in which repetitive periodic cessation of breathing for 10 seconds or more occurs during sleep despite increased effort to breathe. It leads to day-time sleepiness, poorer health, increased healthcare and higher work-related and road accidents costing the national economy billions of dollars per year. Early intervention may improve health outcomes for the sufferers. In this article, a hierarchical diagnostic approach is proposed in which at first a quick and safe three-dimensional (3D) surface imaging based technique is used to identify patients susceptible to OSA, thereby allowing a cost-effective patient screening. The susceptible patients are referred for volume imaging such as Cone Beam Computed Tomography (CBCT) from which the airway and other hard-tissue anatomical features can be extracted. Age and gender specific 3D facial norms and different thresholds have been proposed to compute against which individualized features can be judged to determine the presence of OSA. Finally, the severity of OSA is measured by polysomnography sleep study only for those patients who are confirmed for OSA by both surface and volume image-based analysis.

1 Introduction

Sleep apnoea is a serious health issue with significant public health implications [6, 13]. There are three types of sleep apnoea: obstructive (OSA), central (CSA) and mixed (combination of the two). In OSA (84% of cases), mechanical factors play an integral role in the reduction of airflow despite continued respiratory effort [1]. In CSA (0.4% of cases) the physiological respiratory control processes fail to maintain the required respiratory function for optimal health.

OSA is characterised by the presence of apnoeas (i.e. a complete cessation of breathing despite respiratory effort) or hypopneas, defined as greater than 30% reduction in chest and/or abdominal expansion during breathing or shallow breathing lasting at least 10 seconds combined with at least a 4% reduction in oxygen desaturation. Numerous indices have been developed to express the severity of sleep apnoea diagnosed using polysomnography and include the apnoea index (AI) which represents the total number of apnoeas per hour and the Apnoea-Hypopnea Index (AHI), which represents the total combined apnoeas and hypopnoeas per hour. The AHI has been divided into severity scales: mild ($5 < \text{AHI} < 15$), moderate ($15 < \text{AHI} < 30$) and severe ($\text{AHI} > 30$). Additional indices that have been utilised include sleep arousals (Respiratory Disturbance Index) and subjective patient perceptions of sleep impact on daytime activities (Epworth Sleepiness Scale).

During apnoeic episodes, arterial blood oxygen saturation decreases, and sympathetic activity and blood pressure increases. Each apnoeic episode ends with an arousal from sleep, resulting in marked fragmentation of sleep in affected individuals. Excessive daytime sleepiness is a major consequence of OSA. OSA has also been linked to significant conditions such as hypertension [16, 9], ischaemic heart disease and stroke [18], premature death [17], and impairment of cognitive functions [8] which may contribute to motor vehicle and workplace related accidents (comparable to functioning while intoxicated) [7]. A study from The University of British Columbia demonstrated that a person with OSA is twice as likely to be involved in a motor vehicle accident [19]. For untreated individuals, it has been established that there is a 37% higher 5-year morbidity and mortality rate [14].

It is estimated that 775,000 Australians (4.7% of the adult population) suffer from OSA [15]. The Busselton (Australia) Health Survey [2] of 294 men aged 40 to 65 years revealed that about 26% of individuals have mild and 10% have severe levels of sleep apnoea. The total financial and non-financial burden of OSA in Australia was estimated as 21.2 billion dollars in 2010 including direct health care cost of $575.42 million and indirect health care cost (due to lost productivity, deadweight loss, workplace/motor vehicle accidents, social security payments etc.) of $2.6 billion [18]. In U.S. it was estimated in 2008 that the average additional annual health care cost of an untreated sleep apnoea patient is US $1,336 contributing an estimated total of $3.4 billion/year additional medical costs [1].

In this article, we introduce a novel quantitative diagnostic method for OSA based on the combination of two approaches related to two different imaging modalities (surface and volume). The first approach is based on the analysis of a three-dimensional surface scan of a subject (using e.g. a 3dMD face scanner). We propose to extract quantitative facial features from the scan to differentiate between facial morphologies of OSA patients and normal non-apnoeic individuals. The relative position of the upper and lower jaws to the skull base and in turn to each other can be assessed as represented by the external facial appearance. These facial features can be evaluated to determine the relationship between facial morphology and the severity of OSA. 3D surface facial scanning has the advantage of being a non-invasive imaging tool which does not require exposure to ionizing radiation. The second approach relates to the application of state-of-the-art dental imaging in the form of a Cone Beam CT to obtain a 3D (volumetric) representation of the hard and soft tissues. The determination of the morphology (shape and structure) of the airway of OSA patients should help in revealing any significant deviations from the airway of normal individuals. As Cone Beam CT is a readily available imaging tool in most clinics, the proposed diagnostic method is easily accessible with many control non-OSA patients imaged for unrelated dental anomalies. The overall outcome of the article is the development of improved conservative diagnostic methods which will be accessible to wider patient groups and will contribute in early intervention.

The rest of the article is organized as follows. Various approaches currently used for the diagnosis of OSA is described in Section 2. The conceptual framework for our proposed approach is elaborated in Section 3. Proposal for the evaluation of the new diagnostic method is discussed in Section 4 followed by the conclusions in Section 5.

2 Existing Diagnostic Approaches for OSA

OSA is seen more frequently in older males and is related to many predisposing factors such as increased Body Mass Index (BMI), increased neck circumference, smoking, alcohol consumption and enlarged tonsils and adenoids. Clinicians also recognise specific dentofacial deformities which predispose individuals to the development of OSA. The obvious retrusion or underdevelopment of the lower jaw and and/or the upper jaw alerts the clinician to the possibility of a patient susceptible to OSA.

Today, overnight polysomnography remains the 'gold standard' diagnostic method for OSA. It is a monitored sleep study to record biophysiological changes that occur during sleep. Measurements include electroencephalogram, electrooculograms, submental electromyogram, oronasal airflow, chest wall motion, and arterial oxygen saturation. In addition to the significant inconvenience to the patient, polysomnography requires sophisticated specialist facilities, technical and scientific staff and sleep clinicians, which are commonly not available in all regions.

Imaging techniques have been considered as useful adjunctive tools to diagnose and plan the treatment of OSA, with the radiographic head film (cephalometric) analysis being the most convenient and widely used [3]. However, the cephalometric analysis is inherently limited because of its two dimensional imaging and the lack of information about the airway volume and dimensions [5]. In addition, measurements are obtained with the patient in the upright position which may not accurately reflect the distortion of the airway in the supine sleeping position. This may create an underestimation of the degree and pattern of airway narrowing and/or collapse. Lee et al. [10, 11, 12] analysed facial characteristics to predict OSA with an accuracy of 76.1% using 2D photographic and cephalometric images. These have limitations compared to 3D surface and volume data. For example, while they demonstrated a relationship between facial structural measurements such as alar width and intercanthal distance, they did not assess 3D positional relationships of the relevant structural components representing the underlying jaw base, which is the focus of this article.

During the last few years, there has been significant interest in developing conservative, cost-effective, patient-convenient and widely applicable methods to diagnose and treat OSA. Although the morphology of patients diagnosed with OSA has been well documented using two dimensional (2D) imaging techniques, and to a much lesser degree using 3D imaging techniques, no specific stratified evaluation has demonstrated the impact of progressive distortions of the maxillomandibular structures on airflow and sleep performance.

3 Proposed Methods and Techniques

Considering the cost effectiveness and the simplicity, we propose a hierarchical framework for diagnosing OSA. We would like to keep the cheaper and widely accessible measures at the beginning and thus screening out a number of patients before suggesting for more expensive and exhaustive approaches. The detailed framework is described in this section.

3.1 Statistical Design

A null hypothesis for developing the new diagnostic approach can be defined as follows: there will be a statistically significant difference in the proportion of patients who are correctly diagnosed with OSA using the new method as compared to the gold standard.

The sample size for the above hypothesis can conservatively be estimated using an expected sensitivity (probability of correctly identifying a patient as positive by the proposed approach given they have OSA) of 0.85 and specificity (probability of correctly identifying a patient as negative by the new approach given they do not have OSA) of 0.95, and a 95% confidence level. A sample size of 100 OSA patients and 100 non-OSA participants would provide a 0.07 precision for sensitivity and 0.04 precision for specificity.

3.2 Determination of Norms and Thresholds

This approach requires a prior set up of age and gender specific facial norms (nn) used as references. For that purpose, we propose to compute the age and gender specific average faces from a large sample of non-OSA subjects. In addition to these average-face norms, we also propose to determine some other thresholds associated with other discriminating features as illustrated in Fig. 1 and explained below.

Threshold t_1 can be established as follows from 3D ear to ear facial surface images (e.g. Fig. 2) of the 100 patients diagnosed with OSA by polysomnography. The face area can be detected and cropped and various surface features (e.g. length of the maxilla, mandible and chin and the circumference of the neck) can be extracted. The relative shape ratios (RSRs) of these different features (e.g. length of maxilla with respect to the mandible and that of maxilla and mandible compared to the forehead and neck) can be computed. These features then can

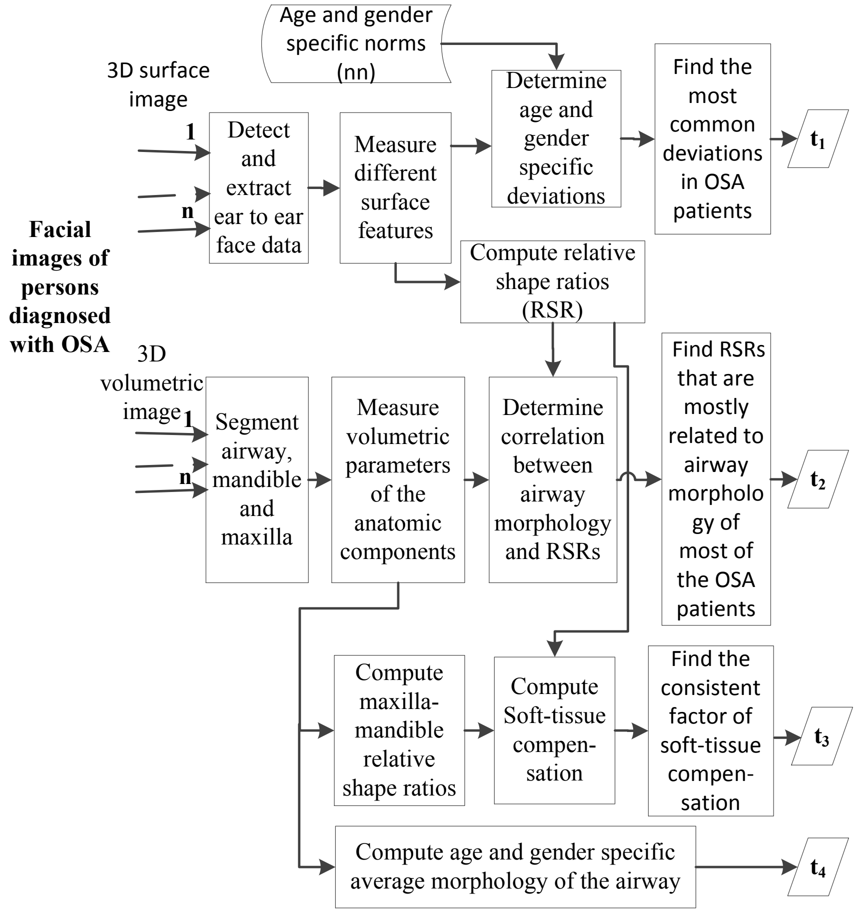


Fig. 1. Block diagram of the computation of different thresholds (t_1, t_2, t_3, and t_4) used in the proposed diagnostic algorithm.

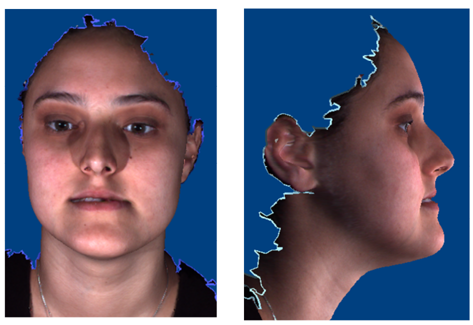


Fig. 2. 3D textured image of a person's frontal (left) and right profile.

be compared with the age and gender specific norms to outline any deviations from the norms. The threshold t_1 can then be derived from these deviations.

Three more thresholds can be determined from 3D volumetric images which can be acquired using a Cone Beam CT scanner from the same patients above. The volumetric data of the airway (Fig. 3) and other anatomical features can be segmented from these data using commercial software such as Dolphin, 3dMD-vultus and 3D Slicer. Different volumetric parameters can be measured and statistically correlated with the facial RSRs computed from the facial surface images. The RSR (of each age and gender group) with the highest correlation factor can be used as a threshold (t_2). The relative shape ratio of maxilla and mandible computed from volumetric data can be compared with those obtained from surface data (3dMD) to evaluate the most common soft-tissue compensation factor (t_3). The average morphology of the airway (threshold, t_4) of the different age and gender subgroups can be computed using the above software or computer programming using MATLAB.

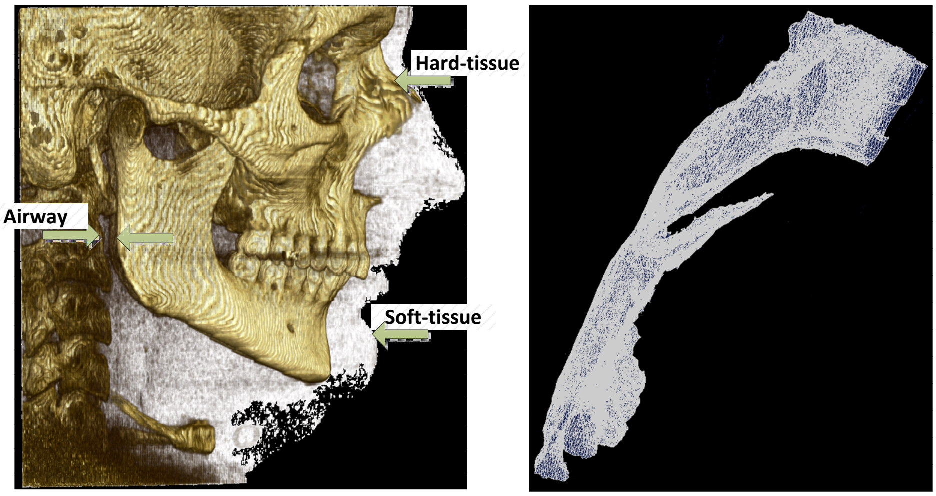


Fig. 3. 3D volumetric image of an OSA patient and his digitally segmented airway represented in wireframe model.

3.3 Diagnosis Using the New Approach

As illustrated in Fig. 4, in the proposed diagnostic framework, a subject presenting for an OSA test will firstly be diagnosed using a surface image. A 3dMD scan (e.g. Fig. 2) will be taken using the 3dMD Facial Scan System. The captured image data will be represented as a 3D surface mesh. Then quantitative facial shape features and ratios will be extracted or derived from the surface data.

An individualized norm will be determined based on the age and gender specific norms (nn) to localize and quantify any shape deviations (d_1) of the facial

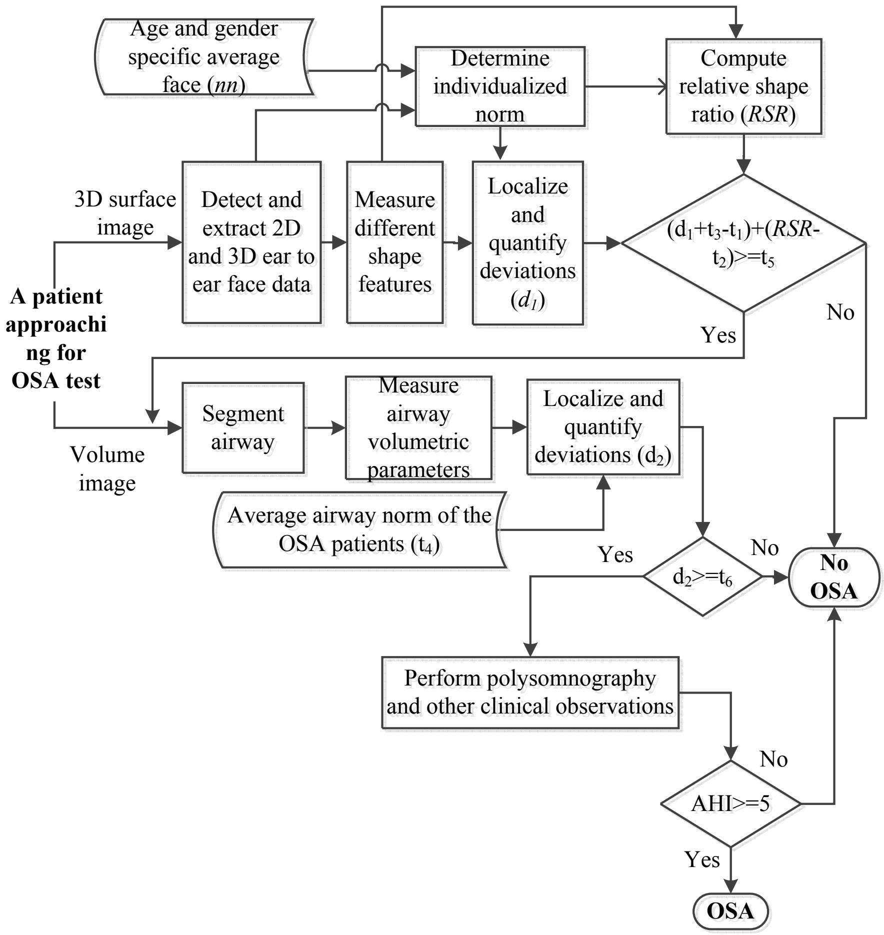


Fig. 4. Block diagram of the proposed diagnostic methods.

shape of the patient relative to a non-OSA subject of the same gender and age group. These shape deviations along with a soft-tissue compensation factor (t_3), will be compared with the morphological threshold t_1. Furthermore, the RSR of the patient will also be compared with the threshold t_2. After computation of these thresholds and deviations obtained from the analysis of the surface image only, patients will be primary identified as a candidate for OSA if the summation of the following two differences is greater than or equal to an empirically evaluated threshold t_5: (i) the difference of the soft-tissue deviations including any soft-tissue compensations from the most common deviations in OSA patients and, (ii) the difference of patients' relative shape ratio from the similar ratio of the most of the OSA patients. The condition can be mathematically represented as in Equation 1.

$$(d_1 + t_3 - t_1) + (RSR - t_2) >= t_5 \tag{1}$$

The potential subjects will then be exposed to a Cone Beam CT scan for a segmental airway assessment, and volumetric parameters of the airway will be measured. Comparing the average airway norm of OSA patients (t_4), the deviation (d_2) in the airway will be calculated. If the subjects' deviations are greater than or equal to an empirically determined threshold t_6, they will be recommended for a polysomnographic sleep study and other clinical observations in order to finally confirm the presence and severity of OSA expressed in AHI.

4 Evaluation of the Proposed Diagnostic Method

A comparison of the new 3D imaging-based diagnostic method with findings from polysomnography can be performed through a test for difference in proportions for the paired-sample design [4]. The diagnosis can be defined as successful if AHI (found using polysomnographic sleep study) of the positively diagnosed patients (using the proposed method) is found to be greater than 5 (the threshold measure of apnoea).

The test for difference in proportions for the paired-sample design can then be used to reject the hypothesis that there will be a statistically significant difference in the proportion of patients who are correctly diagnosed with OSA using the proposed method as compared to the gold standard (polysomnography). If there is no statistically significant difference in the proportion of patients who benefit from the proposed 3D image-based approach, then it should be widely adopted. The test can be specifically described as follows:

1. For the 200 randomized subjects, apply the 3D imaging-based diagnosis approach (response Y1) and standard polysomnography procedure (matched control, response Y2) [4].
2. Define a failure by a 'miss' and 'false alarm' (adopting the terminology of detection theory), i.e. if the subject is diagnosed with the proposed approach while they are not diagnosed using polysomnography, then it is a false alarm. We then determine the proportion of cases when the proposed method resulted in a success (P1) and when polysomnography resulted in success (P2).

3. If the proportions P1 and P2 computed above are equal, then reject the hypothesis that the proportion of successes is the same for our 3D imaging-based diagnosis method and polysomnography. (test statistics is unit-normally distributed; the exact formula is given in [4]).

5 Conclusions

The proposed conservative, cost-effective, patient-convenient and widely applicable methods to diagnose OSA will facilitate more accessible diagnosis of a larger number of patient groups than is possible with polysomnography and will enhance early intervention.

The purpose of the proposed diagnostic approach is not to replace the sleep studies but to screen and then to stratify adult OSA patients for various modes of treatment based on the anatomical features and airflow measurements. Importantly, the proposed approach will also provide guidance to clinicians who manage significant jaw structure problems in children with occasionally irreversible conventional orthodontics with little regard for the consequences of leaving the child prone to developing sleep apnoea with their underlying jaw structure remaining atypical. The specific patterns of jaw morphology can be identified during the diagnosis in individuals who would be considered for surgical management of their jaw deformity in adulthood rather than attempting to compensate the teeth for the jaw structure. Clinicians may then modify the way in which they advise patients with more severe jaw structure problems based on the impact on predisposition to OSA. Moreover, after screening, morphologically predisposed patients may be warned about lifestyle habits which may contribute to the possibility of developing OSA at a later age.

Acknowledgments

This research is sponsored by two Special Donation grants from the Australian Society of Orthodontists Foundation of Research and Education (PG51311900 and PG51312000) and, by a Research Development Award (PG12104373) and a Research Collaboration Award (PG12105002) from the University of Western Australia.

Bibliography

[1] Aarab G, Lobbezoo F, Hamburger H, Naeije M (2010) Effects of an oral appliance with different mandibular protrusion positions at a constant vertical dimension on obstructive sleep apnea. Clinical Oral Investigations 14(3)

[2] Bearpark H, Elliott L, Grunstein R, et al (1995) Snoring and sleep apnea. a population study in australian men. Am J Respir Crit Care Med 151:1459–1465

[3] Conley R (2011) Evidence for dental and dental specialty treatment of obstructive sleep apnoea. part 1: the adult osa patient and part 2: the paediatric and adolescent patient. J Oral Rehabil 38(2):136–56

[4] Connor R (1987) Sample size for testing differences in proportions for the paired-sample design. Biometrics 43(11):207–211

[5] Doff M, Hoekema A, Pruim G, van der Hoeven J, de Bont L, Stegenga B (2009) Effects of a mandibular advancement device on the upper airway morphology: a cephalometric analysis. J Oral Rehabil 36(5):330–7

[6] George C (2001) Reduction in motor vehicle collisions following treatment of sleep apnoea with nasal cpap. Thorax 56(7):508–512, DOI 10.1136/thorax.56.7.508

[7] Horne J, Reyner L (1995) Sleep related vehicle accidents. Bmj 310(6979):565–7

[8] Klitzman D, Miller A (1994) Obstructive sleep apnea syndrome: complications and sequelae. Mt Sinai J Med 61(2):113–21

[9] Lavie P, Herer P, Hoffstein V (2008) Obstructive sleep apnoea syndrome as a risk factor for hypertension: population study. Bmj 320(7233):479–82

[10] Lee R, Chan A, Grunstein R, Cistulli P (2009) Craniofacial phenotyping in obstructive sleep apnea - a novel quantitative photographic approach. Sleep 32(1):37–45

[11] Lee R, Petocz P, Prvan T, Chan A, Grunstein R, Cistulli P (2009) Prediction of obstructive sleep apnea with craniofacial photographic analysis. Sleep 32(1):46–52

[12] Lee R, Sutherland K, Chan A, Grunstein R, Cistulli P (2010) Relationship between surface facial dimensions and upper airway structures in obstructive sleep apnea. Sleep 33(9):1249–1254

[13] Lee W, Nagubadi S, Kryger M, Mokhlesi B (2008) Epidemiology of obstructive sleep apnea: a population-based perspective. Expert Rev Respir Med 2(3):349–364

[14] Marti S, Sampol G, Munoz X, Torres F, Roca A, Lloberes P, et al (2002) Mortality in severe sleep apnoea/hypopnoea syndrome patients: impact of treatment. Eur Respir J 20(6):1511–8

[15] Morgenthaler T, Kagramanov V, Hanak V, et al (2006) Complex sleep apnea syndrome: is it a unique clinical syndrome? Sleep 29:1203–1209

[16] Nieto F, Young T, Lind B, Shahar E, Samet J, Redline S, et al (2000) Obstructive sleep apnoea syndrome as a risk factor for hypertension: population study. Jama 283(14):1829–36

[17] Rapoport D (1994) Treatment of sleep apnea syndromes. Mt Sinai J Med 61(2):123–30

[18] Report DAE (2011) Re-awakening australia - the economic cost of sleep disorders in australia. URL http://www.sleephealthfoundation.org.au/research.html

[19] ScienceDaily (2008) Sleep apnea doubles car crash risk, study shows. retrieved on oct 30, 2012. URL http://http://www.sciencedaily.com/releases/2008/02/080218214401.htm