Obstructive Sleep Apnea Stage Classification from Single Lead ECG Using Random Forest Classifier

Md Juber Rahman
The University of Memphis
mrahman8@memphis.edu

ABSTRACT
In this study, we have investigated the use of heart rate variability (HRV) features extracted from single lead electrocardiogram (ECG) signal in the stage classification of obstructive sleep apnea (OSA). We have explored the time domain and frequency domain heart rate variability features, made a statistical analysis and finally classified patients suffering from sleep apnea using a Random Forest Classifier into severe and non-severe group. Results indicate that using only HRV features severe OSA can be classified with a precision of 77.3%, recall of 100% and F score of 87.2%. This reveals the possibility of severe OSA detection as well as monitoring during the treatment period at home environment using smart watch or smartphone enabled with ECG sensor.

Keywords
Single lead ECG, Obstructive Sleep apnea, bagging tree, Random Forest, Classification

INTRODUCTION
Sleep apnea is a very common disease and prevail all over the world. Specially obese and elderly people suffer from sleep apnea. Sleep apnea is responsible for many health problems including insomnia, depression, daytime drowsiness, hypertension etc. in the initial stage and memory loss, stroke, heart attack etc. at the later stage. Sleep apnea can be of two types- obstructive sleep apnea (OSA) and central sleep apnea. OSA is more prevalent and can be treated with continuous positive airway pressure (CPAP) device at home environment. American Academy of Sleep Medicine (AASM) defined OSA as a sleep event of paused breathing where the volume of air entering the lungs drops below 50%, oxygen saturation drops by at least 4% and the event persist for at least 10 sec. The gold standard for diagnosing OSA is polysomnogram.

Polysomnogram is a test that is conducted in a sleep laboratory and collects ECG, EEG, EMG, Nasal airflow, Respiration signal, Oxygen saturation level etc. Based on all these tests OSA is scored by a trained sleep technician. The scale for scoring OSA is called Apnea-hypopnea index (AHI). AHI is defined as the no. of apnea hypopnea event per hour. The gold standard of sleep stage classification is based on AHI. If the AHI< 5 then OSA is classified as normal, if 5< AHI<15 then OSA is classified as mild, if 15<AHI<30 then OSA is classified as moderate and for AHI>30, OSA is classified as severe. AHI is calculated based on the polysomnogram test. Since, no. of sleep laboratory is few and the polysomnogram test is not user friendly, many OSA patients avoid this test and remain undiagnosed and untreated for lifetime. This not only affects his personal health but also sometimes create societal problems such as sleep deprived partner that may lead to even divorce, motor accident due to excessive drowsiness, mouth breathing that leads to bad breath and hence social communication capability reduces significantly etc. Again, it is not possible to conduct polysomnogram test at home environment, hence an OSA patient who is taking treatment with CPAP device can't monitor whether his condition is improving or not.

While the importance of polysomnogram in no way can be undermined for accurate diagnosis of OSA severity, there have been an ongoing effort by researchers to classify sleep apnea based on non-polysomnographic measures. Kumar et al. investigated the use of Gabor filter based one dimensional local phase descriptors for OSA

detection using single lead ECG [1]. Song et al. investigated the detection of OSA using a Hidden Markov model from ECG signals [2]. Chen et al. proposed an automatic screening approach for OSA diagnosis based on single lead electrocardiogram [3]. Although, there have been good progress in classifying apneic patients from healthy subjects based on non-polysomnographic measures, the effort to classify severe OSA patients from the mild and moderate stage patients have been started very recently. Jung et al. approached real-time automatic apnea event detection and estimation of AHI using nocturnal pulse oximetry [4]. In this study, we have investigated the stage classification of OSA from single lead ECG only. We have considered both mild and moderate OSA in the non-severe stage and severe OSA in the severe stage. This algorithm is computationally less intensive and highly suitable for implementation on smartphones for real-time OSA monitoring during CPAP treatment or in the primary detection of severe OSA before undergoing a confirmative polysomnogram test in a designated sleep laboratory.

1. DATASET

The apnea-ECG database is an open access database for development and evaluation of ECG-based apnea detection [5]. The database has 70 records in total out of which 20 records are from healthy subjects, 10 from normal OSA category and the rest of them are in mild, moderate or severe stage. For training and testing purpose, I have divided the dataset in three parts, training set and testing set. In the training set the no. of records are 52 and in the testing set the no. of records are 18. The dataset contains header file, beat annotations, apnea annotations and the digitized ECGs. The sampling rate of originally collected signal is 100 samples/second and the length of each record is approximately 8 hours. All the records have been collected during overnight sleep test in a sleep laboratory using medical grade single lead ECG machine. The electrode position was modified lead V2.

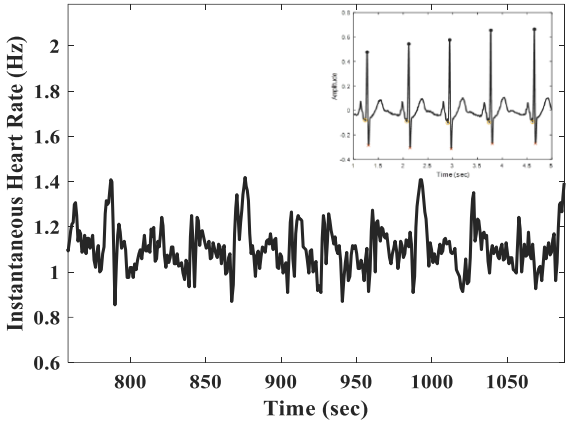

Fig.1 Instantaneous heart rate and ECG signal (Inset)

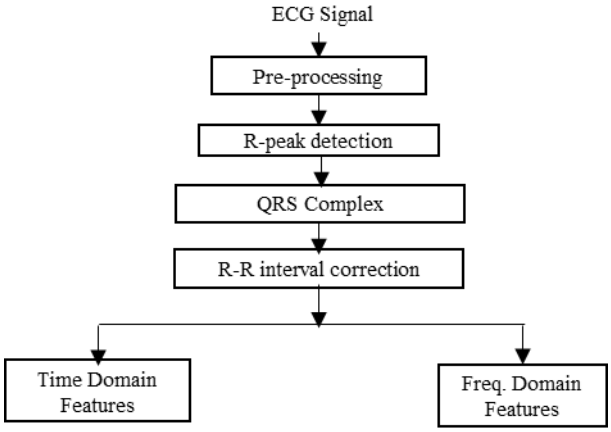

Fig. 2 Feature Extraction Process

2. METHODS

In this section, I have discussed the method of feature extraction and classification including a statistical analysis.

2.1 Feature Extraction

Time domain and frequency domain features have been extracted from the ECG signal. The extracted features are well defined in the literature and are recommended for use in OSA detection [6]. The detailed procedure or feature extraction have been described in Fig.1 .The raw ECG signal is available in .dat format which has been converted into .mat format in Physiobank ATM. and also has been preprocessed which includes low pass and high pass filtering, baseline wander removal etc. In the preprocessing stage the ECG signal has been resampled as 200 Hz. Then Pan-Tompkin algorithm has been used for R-peak detection and also QRS morphology

detection. NN interval has been calculated from R-R peak by removing premature beats. From NN interval time domain and frequency domain features have been extracted. The name and definition of the features have been listed in Table 1. and Table 2. In total 14 features have been extracted. MATLAB software has been used in signal processing and feature extraction. After that the feature table has been exported in a csv file for using in the statistical analysis and classification process.

Table 1. Time Domain Features from ECG Signal

Feature	Description
AVNN	Mean of NN-interval
SDNN	Standard deviation of all NN intervals.
SDANN	Standard deviation of the averages of NN intervals in all 5 min segments of the entire recording.
RMSSD	Defined as the square root of the mean of the squares of differences between adjacent NN intervals.
SDNN index	Mean of the standard deviation of all NN intervals for all 5 min segments of the entire recording.
SDSD	Defined as the standard deviation of the differences between adjacent R-R intervals.
pNN50	NN50 count divided by the total no. of all NN intervals where NN50 is the number of pairs of adjacent NN intervals differing by more than 50 ms.

Table 2. Time Domain Features from ECG Signal

Feature	*Units*	*Description*	*Frequency Range (Hz)*
VLF	s^2	Power in very low frequency range	≤ 0.04
LF	s^2	Power in low frequency range	0.04 – 0.15
HF	s^2	Power in low frequency range	0.15 – 0.4
Total Power	s^2	Variance of all NN intervals	≤ 0.4
LFnu	n.u.	LF power in normalized units	
HFnu	n.u.	HF power in normalized units	
LF/HF		Ratio of LF/HF	

2.2 Statistical Analysis

In the medical community, there is always a prejudice against machine learning techniques and it is advised to do a statistical analysis before applying machine learning techniques. If in the statistical analysis a significant difference is found between the two classes/groups for the features to be used in machine learning classifer then classification using those features make more sense and is readily accepted by physicians and medical researchers. Because there is always a chance for a machine learning method to learn a dataset and over fit and give a good performance

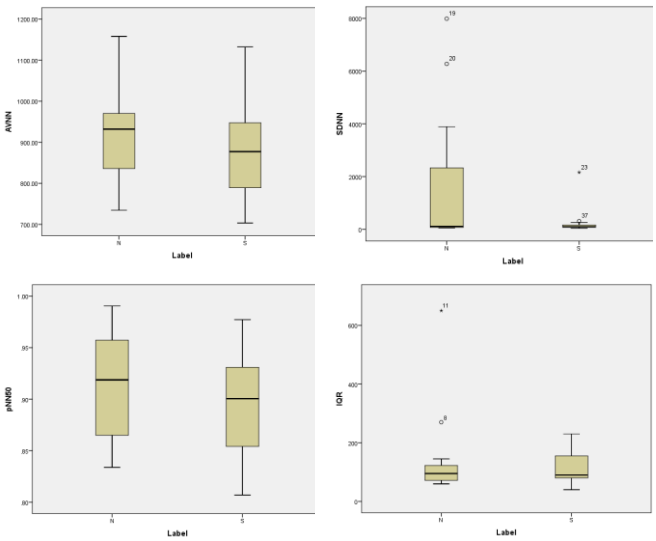

Fig. 2 Box plot of AVNN, SDNN, pNN50, IQR in the severe and non-severe OSA group

Table 3. Result of Welch's t test

Variable	P- value	Decision
AVNN	0.288	Not Significant
SDNN	**0.025**	**Significant**
rMSSD	**0.025**	**Significant**
SDSD	**0.025**	**Significant**
pNN50	0.291	Not Significant
IQR	0.655	Not Significant
Dminmax	**0.026**	**Significant**
SDANN	**0.021**	**Significant**
SDNNindex	0.037	**Significant**
HF	0.282	Not Significant
LF	0.093	Not Significant
Total Pwr	0.229	Not Significant
LFnu	**0.002**	**Significant**
HFnu	**0.002**	**Significant**

in the test set especially if there is any pattern in the test set and the samples are not collected at random and the size of test set is small. That's why a statistical analysis have been done to confirm the suitability of the features. Box plot have been used to compare the two groups and Welch's *t* test test has been used to see if the difference between the two groups are statistically significant. The statistical software SPSS has been used for statistical analysis. While the boxplot in Fig.2 shows a difference between the mean of two groups the *t* test result in Table 3 confirms that the difference is statistically significant for most of the features (statistically significant features have been bold marked).

2.3 Classification

Since the dataset is labeled, supervised learning method has been adopted for solving the classification problem. Initially, the dataset has been divided in training and test set where the training set has 52 records and test set has 18 records. Five-fold cross validation has been used in the training set to check the predictive performance of the classifier and parameter tuning to achieve the sweet spot and avoid under fitting/overfitting. Weka has been used as the classification environment. Several classifier have been tested, out of which Random Forest classifier performed best. Random forest is an ensemble learning algorithm which uses decision trees as the base classifier and a bagging method of learning [7]. Ensembles are a divide and conquer approach used to improve performance. The main principle behind ensemble methods is that a group of weak learners can come together to form a strong learner. In simple words, Random forest builds multiple decision trees and merge them together to get a more accurate and stable prediction. In addition, it brings extra randomness into the model when it is growing the trees. The tuned Random Forest classifier had batch size of 100, bagSizePercent 100, unlimited maximum depth, and one execution slot.

3. RESULTS

The five-fold cross validation performance of different classifiers has been shown in Table 4. It is clearly evident that Random Forest is the best performer in terms of precision, recall and F1-score. The ROC area was 0.752 and PRC area 0.904. Time taken to build the model was 0.02 second.

Table 3. Result of Five-fold cross validation

Classifier	Precision	Recall	F-score
Support Vector Machine	0.708	1.00	0.829
Adaboosting	0.727	0.941	0.821
Random forest	0.773	1.00	0.872

Table 3. Classification Performance in the test set

Classifier	Precision	Recall	F-score
Random forest	0.839	0.892	.864

4. CONCLUSION AND DISCUSSIONS

Recent advances in sensor technology and machine learning techniques has paved the way for using smartphone in the reliable monitoring and detection of OSA. This study reveals that using heart rate variability features in the classification of OSA stages provides a reasonable accuracy which can be further improved with feature selection and adding other non-polysomnographic features such as pulse-oximetry. In future, deep learning based technique will be used to obtain a better accuracy and implementation of this model in android environment for real-time usage will be explored.

5. ACKNOWLEDGMENT

I would like to thank Dr. Deepak Venugopal for offering the machine Learning course and helping me understand the core concepts.

REFERENCES

[1] T. S. Kumar and V. Kanhangad, "Gabor Filter-Based One-Dimensional Local Phase Descriptors for Obstructive Sleep Apnea Detection Using Single-Lead ECG," in *IEEE Sensors Letters*, vol. 2, no. 1, pp. 1-4, March 2018. doi: 10.1109/LSENS.2018.2807584

[2] C. Song, K. Liu, X. Zhang, L. Chen and X. Xian, "An Obstructive Sleep Apnea Detection Approach Using a Discriminative Hidden Markov Model From ECG Signals," in *IEEE Transactions on Biomedical Engineering*, vol. 63, no.7, pp.1532-1542, July2016.

doi: 10.1109/TBME.2015.2498199

[3] L. Chen, X. Zhang and C. Song, "An Automatic Screening Approach for Obstructive Sleep Apnea Diagnosis Based on Single-Lead Electrocardiogram," in *IEEE Transactions on*

Automation Science and Engineering, vol. 12, no. 1, pp. 106-115, Jan. 2015.doi: 10.1109/TASE.2014.2345667

[4] D. W. Jung *et al*., "Real-Time Automatic Apneic Event Detection Using Nocturnal Pulse Oximetry," in *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 3, pp. 706-712, March2018. doi: 10.1109/TBME.2017.2715405

[5] T Penzel, GB Moody, RG Mark, AL Goldberger, JH Peter. The Apnea-ECG Database .*Computers in Cardiology* 2000; 27: 255-258.

[6] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J* (1996) 17:354–81.10.1093

[7] L. Breiman, "Random Forests", Machine Learning, 45(1), 5-32, 2001