DIAGNOSING SLEEP APNEA VIA FEATURE SELECTION ON SINGLE CHANNEL ECG

Huseyin Guruler[1], Abdullah Ferikoglu[2]

[1]Mugla Sitki Kocman University, Faculty of Technology, Department of Information Systems Engineering, Mugla-Turkey
hguruler@mu.edu.tr
[2]Sakarya University, Faculty of Technology, Department of Electrical & Electronics Engineering, Sakarya-Turkey
af@sakarya.edu.tr
[1]Corresponding Author

ABSTRACT

This article is based on a combination of time-frequency domain functions, and nonlinear techniques in the analysis of heart rate variability (HRV) for diagnosing obstructive sleep apnea (OSA) using only single-lead electrocardiography (ECG) signals. The contribution of the presented study to earlier ones is that it enables numerically determining what type of HRV features better represent the aforementioned target by using correlation matrices and neural networks (NNs).

Keywords: Diagnosing disease, neural network, sleep apnea, heart rate variability, feature selection, correlation matrices

1. INTRODUCTION

Study on variations in the instantaneous heart rate (HR) time series using the beat-to-beat RR intervals are known as HRV analysis. HR increases with sympathetic activity and decreases with parasympathetic (vagal) activity. The balance between the effects of the sympathetic and the parasympathetic systems, the two opposite acting branches of the autonomic nervous system (ANS), is referred to as the sympathovagal balance (SB) (Acharya, Joseph, Kannathal, Lim, & Suri, 2006).

Spectral analysis is typically used to estimate the effect of the sympathetic and parasympathetic modulation of the RR intervals. The two main frequency bands of interest are referred to as the low frequency (LF) band and the high frequency (HF) band. Sympathetic tone is believed to influence the LF component, whereas both sympathetic and parasympathetic activities have an effect on the HF component. The ratio of the power contained in the LF and HF components has been used as a measure of the SB (Jos & Spaan, 2007).

OSA is a serious disorder caused by intermittent airway obstruction in sleep (Abdullah, Maddage, Cosic, & Cvetkovic, 2010), (Lado, et al., 2009). OSA causes changes in cardiac and neuronal activity and discontinuities in sleep pattern when observed via ECG and electroencephalogram (EEG). OSA is usually diagnosed using polysomnography (PSG) conducted in sleep laboratories (Roche, Celle, Pichot, Barthélémy, & Sforza, 2007). PSG is utilized to define physiological sleep and its different stages, to assess sleep quality and to diagnose many types of sleep disorders such as insomnia, OSA, restless legs syndrome and periodic leg movement disorders. However, PSG is very expensive and the technology requires not only the connection of various sensors and electrodes (e.g. EEG, Electrooculogram (EOG) and Electromyogram (EMG), and ECG etc.) but also spending the night in a bed (Yilmaz, Asyali, Arikan, Yetkin, & Ozgen, 2010).

Detection of OSA can be performed and significantly improved through HRV analysis, since fluctuations of oxygen saturation in blood accompanied by apnea, cause variations in the HR (Quiceno-Manrique, Alonso-Hernandez, Travieso-Gonzalez, Ferrer-Ballester, & Castellanos-Dominguez, 2009), (Al-Abed, Manry, Burk, Lucas, & Behbehani, 2009). SB has been used for detection of OSA in many studies. The review (Penzel, et al., 2002), presents systematic comparison of studies using different algorithms for OSA detection based on the same ECG recordings (Moody, Mark, Goldberger, & Penzel, 2000). In these researches, HRV with or without respiratory signals are generally analyzed in three main areas: Time, Frequency and Non-linear Analysis. Each analysis technic produces some features, which are usually numerical values. These values are used in decision-making algorithms or mathematical models that may involve neural networks (NNs), support vector machines (SVMs), wavelet etc. So far, many combinations of time, frequency and non-linear domain features of HRV obtained from ECG have been used with different type of classification methods. Although high accuracies of OSA detection and significant successes on apnea classification can be achieved, it is still unclear which feature parameters are more effective for classification.

This study aims to classify pre-collected sleep data into one of the three basic types; apnea, hypopnea, and healthy episodes, with fewer parameters obtained from single-lead ECG recordings. Besides that, it determines numerically what features of HRV better represent the classification.

2. MATERIALS and METHODS

The present study uses a variety of significant and relevant characteristic features include morphological information, duration and complexity details of the ECG to classification. Table 1. summarize all the HRV measures.

Table 1. Summary of the HRV Measures

Measures		Description
Time-Domain	RR	The mean of RR intervals
	SDNN	Standard deviation of RR intervals
	HR	The mean of HR
	STD HR	Standard deviation of instantaneous HR values
	RMSSD	Square root of the mean squared differences between successive RR
	NN50	Number of successive RR interval pairs that differ more than 50 ms
	pNN50	NN50 divided by the total number of RR intervals
	HRV triangular	The integral of the RR interval histogram divided by the height of the
	TINN	Baseline width of the RR interval histogram
Frequency	Peak frequencies	VLF, LF, and HF band peak frequencies
	Absolute powers	Absolute powers of VLF, LF, and HF bands
	Relative powers	Relative powers of VLF, LF, and HF bands
	Normalized	Powers of LF and HF bands in normalized units
Nonlinear	SD1, SD2	The standard deviation of the Poincare plot perpendicular to (SD1) and
	ApEn	Approximate entropy
	SampEn	Sample entropy
	D2	Correlation dimension
	DFA	Detrended fluctuation analysis
	Alfa 1	Short term fluctuation slope
	Alfa 2	Long term fluctuation slope
	RPA	Recurrence plot analysis
	Lmean	Mean line length
	Lmax	Maximum line length
	REC	Recurrence rate
	DET	Determinism
	ShanEn	Shannon entropy

The feature sets obtained from analysis methods those are time-domain, frequency-domain and non-linear methods. Time domain analysis involved statistical and geometrical calculations. Frequency domain analysis was performed using fast Fourier transform (FFT) and autoregressive (AR) modelings. Non-linear methods including Poincaré and recurrence plots, approximate and sample entropies, detrended fluctuations and correlation dimensions were used.

WFDB (WaveForm Databases) software package used for viewing, analyzing, and creating recordings of physiologic signals. “Kubios HRV” software used for HRV analysis (Niskanen, Tarvainen, Ranta-Aho, & Karjalainen, 2004). Matlab NN toolbox was used for classification.

Classification of OSA was realized on Apnea-ECG Database in PhysioBank (Penzel, Moody, Mark, Goldberger, & Peter, 2000). Table 2 shows Demographic and Clinical Features of the dataset. The data consists of 70 records, divided into learning and test sets equally. Learning and test recordings involves three classes namely Apnea, Hypopnea, and Healthy depending on apnea-hypopnea index (AHI) (Ruehland, et al., 2009).

Table 2. Demographic and clinical features of apnea-ECG dataset

	All subjects	Apnea	Hypopnea	Healthy	*p*-value
Subjects (n)	70	40	10	20	-
Age (years)	45.6±10.6	51.5±7.6	47.2±5.9	32.9±5.4	NS
Males (n)	57	38	8	11	$p<0.01$
BMI (kg/m2)	28.1±6.5	30.8±4.6	30.4±9.2	21.3±1.9	NS
Records (h)	8.2±0.5	8.4±0.4	8.0±0.6	7.9±0.4	NS
AHI (e/h)	-	45.4±22.5	12.1±12.0	0.0±0.0	$p<0.01$

Data are presented as mean ± SD or n; BMI: Body mass index; NS: no significant statistical difference

Depending of AHI: Apnea: Recordings with clear occurrence of sleep apnea (100 min or more). 40 recordings fulfilled this criterion. Hypopnea: Recordings with some degree of sleep apnea (between 5 and 99 min). The recordings revealed either mild apnea, up to an apnea index of 10 events per hour, or obstructive snoring in otherwise healthy subjects. 10 recordings fulfilled this criterion. Healthy: Recordings of healthy subjects with neither sleep apnea (fewer than 5 min) nor habitual snoring. 20 recordings fulfilled this criterion.

Figure 1 describes the classification process.

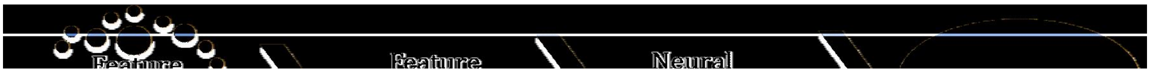

Figure 1. OSA Classification Process

Feature extraction was realized on HRV by time, frequency and non-linear techniques. Due to the feature extraction involves number of parameters having various degrees of importance for classification, CMs were used to select the parameters, which are preferred for neural networks (NNs) as an input. Better correlation provides better classification ability for NNs. CMs are simply tables, in which correlation coefficients "see (1)" for every single column in relation to target column take place.

$$r(X,Y)=\frac{\Sigma\left((x-\bar{x})\right)\left((y-\bar{y})\right)}{\sqrt{\Sigma(x-\bar{x})^2\,\Sigma(y-\bar{y})^2}} \qquad (1)$$

Feed forward back propagation NNs can give high accuracy results with small number of hidden layers. Hence, the classification process was realized with feed forward back propagation NNs.

3. RESULTS

Table 3 and Figure 2 shows the classification and iteration results for 6 methods. Here, only highly correlated parameters in each methods were noticed. CMs helps to find lower correlated parameters, which could be eliminated by looking their correlation coefficients. Methods 1 to 4 indicates classification ability of parameters belonging to each HRV analysis method individually. Methods 5 and 6, on the other hand, involves mixed parameters from methods 1 to 4. Method 6 use only high correlated parameters in its groups.

Table 3. Classification results

Classes		A, B, C	(A+B), C
1	Method	Temporal parameters	
	Accuracy	0.63	0.85
	Iterations	100	60
2	Method	FFT spectral parameters	
	Accuracy	0.78	0.93
	Iterations	500	300
3	Method	AR spectral parameters	
	Accuracy	0.78	0.93
	Iterations	300	70
4	Method	Non-Linear parameters	
	Accuracy	0.72	0.96
	Iterations	400	100
5	Method	All parameters from 1 to 4 methods	
	Accuracy	0.78	0.96
	Iterations	600	50
6	Method	High correlated parameters from 1 to 4 methods	
	Accuracy	0.82	0.96
	Iterations	600	30

*Three Classes are A, B, C and two classes are (A+B), C (A:Apnea B:Hypopnea C:Healthy)

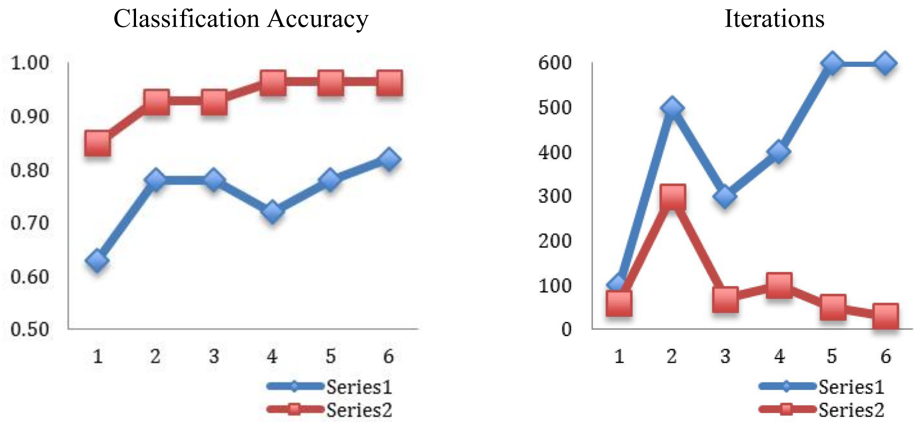

Figure 2. Classification and Iteration Results

Series 1 the results for two classes (A+B) and C, Series 2 the results for three classes A, B and C

Accuracy was high using the parameters of non-linear, frequency and time domain respectively. The accuracy was found 96.42% for the classification to A and C (A- Apnea C- Healthy) and 82% for the classification to A, B and C (A- Apnea B- Hypopnea C- Healthy) respectively. When certain parameters selected only higher correlated from 1 to 4 methods, highest accuracy observed with minimum iteration numbers using NNs.

4. CONCLUSIONS

This work offers a method for feature selection and regularization of classifier parameters that were used to optimize classifier performance. The accuracy of the results with the dimension reduction on the feature sets clearly shows that correlation matrices can be focused on minimizing the feature sets used for sleep classification. Classification performance was highest with non-linear equations and lowest with time domain features, with frequency features at midpoint on OSA diagnosis. On the other hand, computational works were in opposite direction in terms of iterations, which required to obtain reasonable results on NN structures.

5. ACKNOWLEDGMENTS

We would like to acknowledge the funding support for this project from The Scientific and Technological Research Council of Turkey (TÜBİTAK), (2214) International Research Fellowship Program, December 2010-December 2011.

6. REFERENCES

Abdullah, H., Maddage, N. C., Cosic, I., & Cvetkovic, D. (2010). Cross-correlation of EEG frequency bands and heart rate variability for sleep apnoea classification. *Medical & biological engineering & computing, 48*(12), 1261-1269.

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: A review. *Medical and Biological Engineering and Computing, 44*(1), 1031-1051.

Al-Abed, M. A., Manry, M., Burk, J. R., Lucas, E. A., & Behbehani, K. (2009). Sleep disordered breathing detection using heart rate variability and R-peak envelope spectrogram. *IEEE Engineering in Medicine and Biology Society, EMBC 2009*, (pp. 7106-7109).

Jos, S. M., & Spaan, A. E. (2007). *Advances in Cardiac Signal Processing.* Berlin Heidelberg: Springer-Verlag.

Lado, M. J., Vila, X. A., Rodriguez-Linares, L., Mendez, A. J., Olivieri, D. N., & Felix, P. (2009). Detecting Sleep Apnea by Heart Rate Variability Analysis: Assessing the Validity of Databases and Algorithms. *Journal of medical systems, 35*(4), 473-481.

Moody, G. B., Mark, R. G., Goldberger, A. L., & Penzel, T. (2000). Stimulating rapid research advances via focused competition: the computers in cardiology challenge 2000. *IEEE Computers in Cardiology 2000*, (pp. 207-210). Cambridge, MA, USA.

Niskanen, J. P., Tarvainen, M. P., Ranta-Aho, P. O., & Karjalainen, P. A. (2004). Software for advanced HRV analysis. *Computer methods and programs in biomedicine, 76*(1), 73-81.

Penzel, T., McNames, J., Murray, A., de Chazal, P., Moody, G., & Raymond, B. (2002). Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. *Medical and Biological Engineering and Computing, 40*(4), 402-407.

Penzel, T., Moody, G. B., Mark, R. G., Goldberger, A. L., & Peter, J. H. (2000). Apnea-ECG database. *IEEE Computers in Cardiology 2000*, (pp. 255-258). Cambridge, MA, USA.

Quiceno-Manrique, A. F., Alonso-Hernandez, J. B., Travieso-Gonzalez, C. M., Ferrer-Ballester, M. A., & Castellanos-Dominguez, G. (2009). Detection of obstructive sleep apnea in ECG recordings using time-frequency distributions and dynamic features. *IEEE Engineering in Medicine and Biology Society, EMBC 2009*, (pp. 5559-5562).

Roche, F., Celle, S., Pichot, V., Barthélémy, C., & Sforza, E. (2007). Analysis of the interbeat interval increment to detect obstructive sleep apnoea/hypopnoea. *European Respiratory Journal, 29*(6), 1206-1211.

Ruehland, W. R., Rochford, P. D., O'Donoghue, F. J., Pierce, R. J., Singh, P., & Thornton, A. T. (2009). The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. *Sleep, 32*(2), 150-157.

Yilmaz, B., Asyali, M. H., Arikan, E., Yetkin, S., & Ozgen, F. (2010). Sleep stage and obstructive apneaic epoch classification using single-lead ECG. *Biomedical engineering online, 9*, 39.

Huseyin Guruler is academic staff in the Department of Information Systems Engineering in Mugla Sitki Kocman University, Turkey. His bachelor's degree is in the field of electronics and computer from Marmara University. MSc is in the field of statistics and computer in Mugla Sitki Kocman University. PhD is in the field of electronics and computer in Sakarya University. PhD thesis is about diagnosing sleep apnea using ECG signals. His research interests merge in data mining and knowledge discovery besides dealing with computational biology and multi-user computers architecture.