A Time-Series Approach to Predict Obstructive Sleep Apnea (OSA) Episodes

Galip Ozdemir, Huseyin Nasifoglu, Osman Erogul
TOBB University of Economics and Technology
Söğütözü Cad. No: 43 TOBB Ekonomi ve Teknoloji Üniversitesi 06560, Ankara, TURKEY
g.ozdemir@etu.edu.tr, hnasifoglu@etu.edu.tr, erogul@etu.edu.tr

Abstract – Sleep apnea is a common respiratory disorder during sleep. It is characterized by pauses in breathing or shallow breathing during sleep for longer than 10 seconds. Except the fact that not having a proper sleep and being rested for the next day, in some cases the apnea period (not breathing interval) may last more than 30 seconds and this situation can even be fatal. 14% of men and 5% of women suffer from Obstructive Sleep Apnea (OSA) in United States. Patients may experience apnea for more than 300 times in a single night sleep. Polysomnography (PSG) is a multi-parametric recording of biophysiological changes, containing EEG, ECG, SpO$_2$, Nasal Airflow signals, performed during overnight sleep. In this study, a fully automatic apnea detection algorithm is developed and an early warning system is proposed to predict OSA episodes by extracting time-series features of OSA periods and regular respiration using nasal airflow signal. Extracted features are then reduced to improve the performance of the prediction. Support vector machines (SVM), one of the commonly used classification algorithms in medical applications, is implemented for learning and prediction of the OSA episodes. The results show that OSA episodes are predicted with 87.6% of accuracy and 91.3% of sensitivity, 30 seconds before patient faces apnea. By this approach, apnea related health risks can be minimized by foreknowledge.

Keywords: obstructive sleep apnea (OSA), prediction of OSA episodes, nasal airflow signal, support vector machines

1. Introduction

Sleep apnea (SA) is generally expressed as not breathing for at least 10 seconds during the sleep. It is an under-diagnosed sleep disorder and is a fatal risk factor. SA is associated with high risks of hypertension, stroke and also with increased mortality rates [1]. Approximately 14% of men and 5% of woman in United States are suffering from OSA syndrome and incidence rate is increasing worldwide [2]. In a sleep study, the severity level of OSA is measured by the number of apnea and hypopnea events per hour during sleep; known as the apnea-hypopnea index (AHI). A subject having clinical symptoms such as excess daytime sleepiness and impaired cognition in addition to AHI greater than 5, is diagnosed as an OSA patient [3]. AHI is usually calculated through overnight polysomnography (PSG) recorded from suspected OSA patients, in sleep study laboratories. Overnight polysomnography (PSG) is the gold standard method for diagnosis of OSA [4]. PSG requires recordings and monitoring of multi-parameter biophysiological signals, including EEG, ECG, respiratory effort, nasal airflow and oxygen saturation (SpO$_2$). These recorded signals are then analyzed by sleep specialists for final diagnosis of the apnea syndrome. SA is generally characterized by pauses in breathing or shallow breathing when the soft tissue in the rear of the throat collapses and closes during whole night sleep [5]. A slow respiratory rate caused by the blockage in the airway, results with less supply of oxygen delivered from the lungs to heart and body. Without breathing (or shallow breathing), CO$_2$ level in the blood begins to elevate and the patient tries to wake up feeling a choke. There are two commonly faced types for apnea: Obstructive sleep apnea (OSA) and Cental sleep apnea (CSA). In OSA, the airflow through nose stalls for a period of time and but the brain struggles with the body to breath. In CSA case, brain is not able to send stimulus to muscular system to resume breathing. There are two devices to overcome apnea syndrome: continuous positive airway pressure (CPAP) and bilevel positive airway pressure (BiPAP). Both devices provide positive airway pressure during the respiratory cycle. However, the main difference among them is BiPAP applies higher pressures during inspiration than expiration [6].

There are many studies to detect apnea episodes using PSG recordings. Some of them focus on using ECG signals to detect OSA [7,8]. SA is defined as not or shallow breathing during sleep, for this reason it is more accurate to predict apnea by nasal airflow signal. Han et al. proposed a method to detect apneic events based on second derivatives of the respiratory

signal [9]. Some other recent works have focused on apnea syndrome diagnosis, by using nonlinear measures of airflow signal [10].

2. Fully Automatic Sleep Apnea Detection

Sleep apnea is the syndrome of pause in breathing during sleep longer than 10 seconds. Our data set has nasal respiratory airflow signals of 6 patients sampled at 32 Hz for an overnight sleep recorded at a sleep study laboratory in hospital. Each of these recordings has duration about 8 hours. 8-hour long nasal airflow signal is first divided into 40 seconds segments (epochs). Every 40-second signal is evaluated independently in terms of energy by using Teager Energy Operator. Epochs containing potential apnea episodes are determined by changes in the energy level. Then, first order derivative approach is applied to the nasal airflow signal episodes, previously labelled as potential apnea candidate. This approach ensures to determine apnea episode beginning precisely. Figure 1 shows the general schematics of the apnea detection algorithm used in this study.
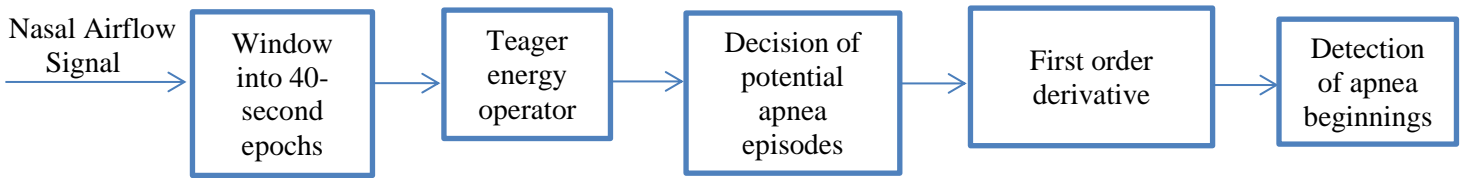
Nasal Airflow Signal → Window into 40-second epochs → Teager energy operator → Decision of potential apnea episodes → First order derivative → Detection of apnea beginnings

Fig. 1: Block diagram of sleep apnea detection.

2.1. Windowing

It is difficult to detect not only a 10-second apnea episode using an 8-hour nasal airflow signal, but also visualize in shorter intervals. For this reason we divided the nasal airflow signal into 40-second epochs. Each epoch is analyzed independently. The purpose of this independent calculation is to minimize the effects of incorrect samples caused by possible movement of the patient or the transducer dislocation during the sleep. Then every epoch is divided into two distinct 20 second parts and the energy of each segment is calculated by Teager Energy Operator.

$$\varphi(n) = x(n)^2 - x(n+1)x(n-1) \tag{1}$$

The energy levels of the two 20 second long signals are compared with respect to total energy of the epoch, whether the ratio of average energies is < 0.25 (or > 4) or not. The segments that have less than one quarter of average energy than the average energy of the epoch are marked as potential apnea episodes. Two potential apnea candidate segments are shown in Figure 2.a and Figure 2.b is an output of the detection algorithm. It can be easily seen that in Figure 2.a the patient could not breathe for longer than 10 seconds. On the other hand, mislabelled epoch is detected as potential apnea episode in Figure 2.b, where there is actually no problem in respiration.
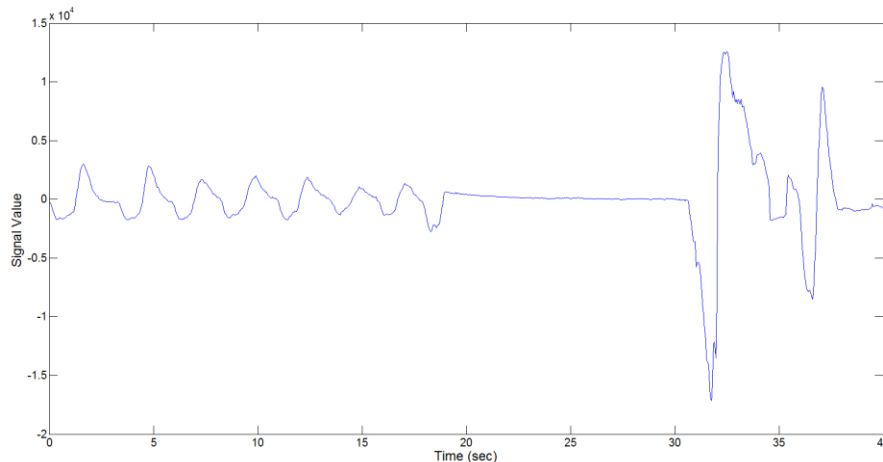

Fig. 2.a: Output of windowing algorithm for an epoch labelled as apnea candidate.

In Figure 2.a, there happens a pause, up to almost 13 seconds on breathing during the second half of the epoch. On the other hand, the total energy of the epoch is four times greater than the energy of one of the 20-second segments.
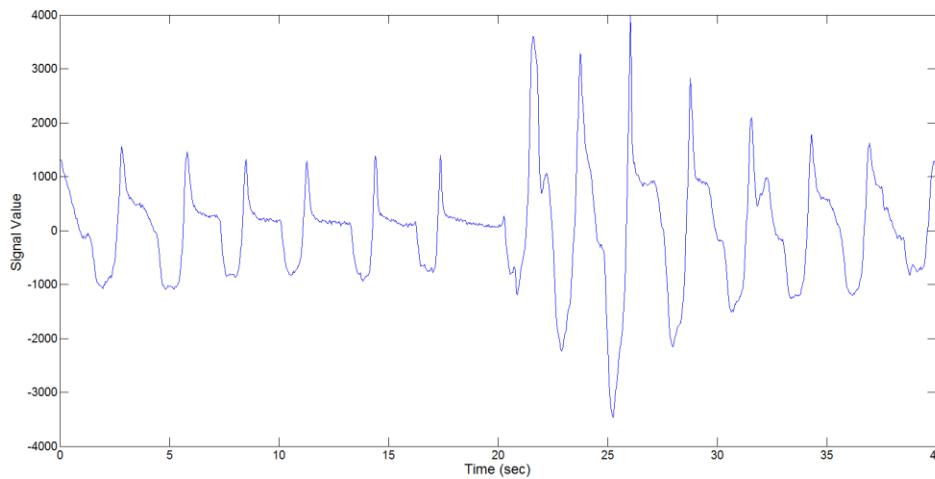

Fig. 2.b: Output of windowing algorithm for an epoch labelled incorrectly as apnea candidate.

In Figure 2.b, the epoch is detected as apnea candidate as there happens four times energy difference between the 20-second segments and the total energy of the epoch. It can be seen that there does not occur any apnea syndrome during the epoch. In order to eliminate the mislabelled regions, first order derivative is applied for the next step.

2.2. First Order Derivative Approach

The aim of first order derivative approach is to eliminate mislabelled epochs like in Figure 2.b. In this part, the derivatives of the epochs are calculated. The first order derivative of every 8-sample interval (0.25 seconds for sampling frequency of 32 Hz) is obtained (Figure 2.c) and compared with a designated threshold value. If the interval length for the derivation is chosen longer, the error in detected apnea beginning precisely increases. Obviously, smaller values may cause other inaccuracies in calculation; considering that nasal airflow signal is not highly deviating. The samples below the threshold value are marked (Figure 2.d) and they are checked whether there is a segment, composed of consecutive samples, that lasts longer than 10 seconds. If not, these samples are removed from the list of the potential apnea episodes. Moreover, the outputs for the epochs which are labelled incorrectly (like an example in Figure 2.b) will also be eliminated because of not containing any 10-second long apnea episodes. In this manner, small deviation values which exist for longer than 10 seconds are detected as apnea event and the initial sample value (marked value) for every apnea periods in the whole signal is stored by examining every potential apnea epochs (Figure 2.e).
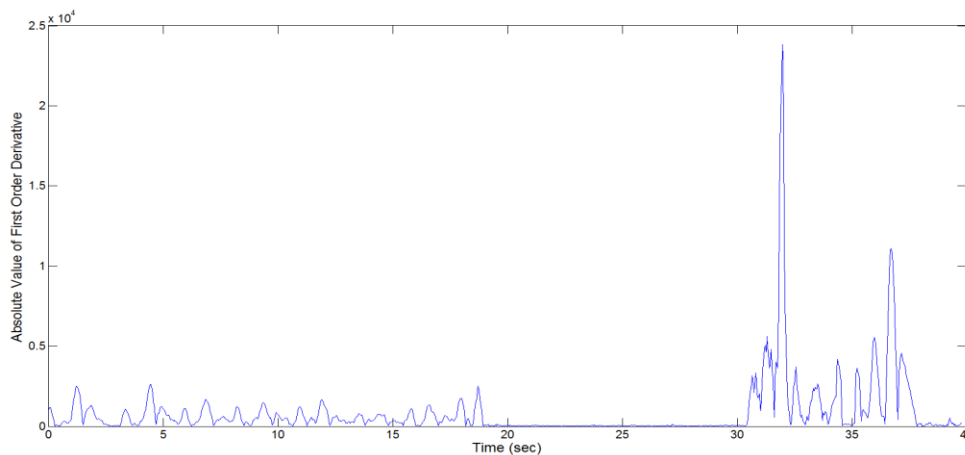

Fig. 2.c: Results for the first order derivative of a potential apnea epoch.
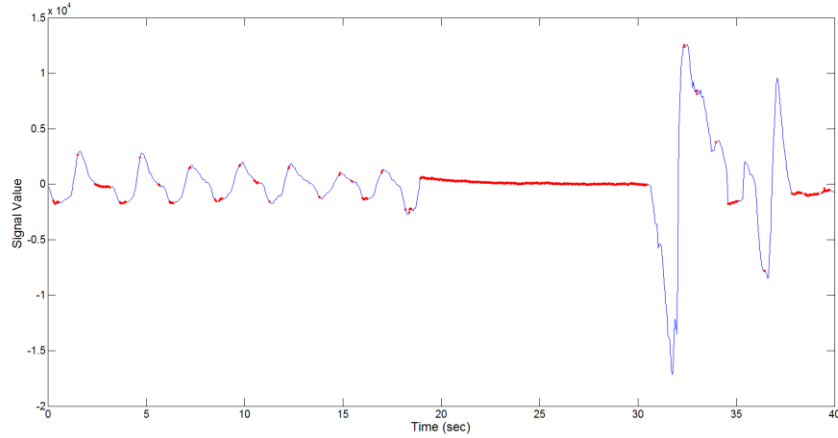
Fig. 2.d: Marked samples after 1st order derivative applied to the nasal airflow signal.
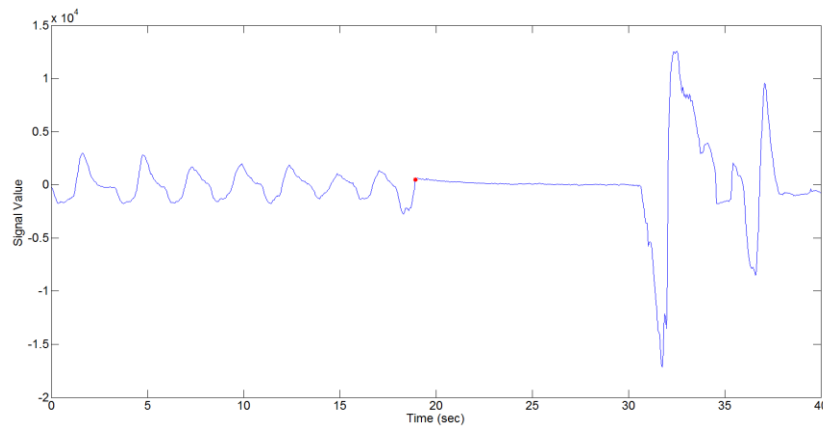

Fig. 2.e: Initial sample of the apnea period is marked after apnea detection.

The output of the first order derivative approach is a single sample value that determines and marks the beginning of the sleep apnea episode. It does not include any other data about the duration of the apnea and it will only be used for distinguishing a time interval labelled as pre-apnea.

After the apnea detection process is accomplished, regular breathing intervals are detected by using similar approach. An example of automatically detected regular air flow (regular breathing) is shown as an example in Figure 2.f.
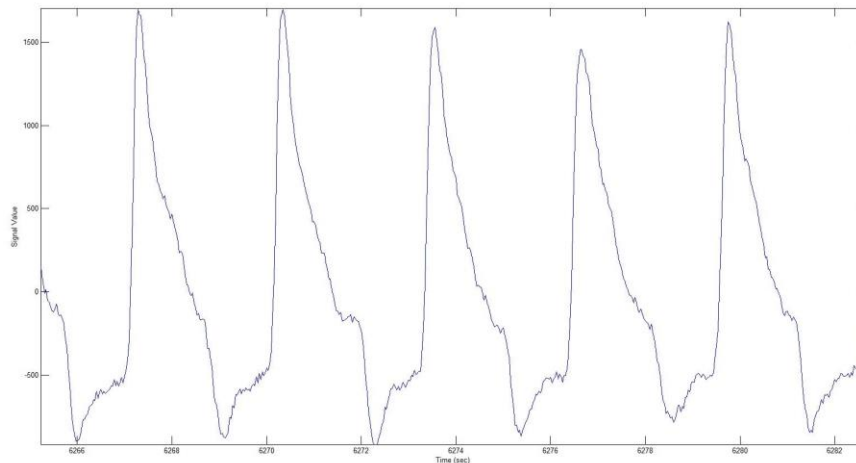

Fig. 2.f: Regular breathing in the airflow signal.

End of this section, OSA segments are automatically detected by the developed algorithm and beginning time of these segments are also marked automatically. Then regular breathing segments are selected from the remaining respiration signal where labelled OSA segments are removed from the original recording. Labelled OSA and regular breathing segments will be used as training inputs to learning algorithm. By this way, the system will be able to predict whether a new signal has an OSA warning or not.

3. Prediction Algorithm

There are two different classes of input, used for prediction: Pre-apnea and regular breathing. Pre-apnea interval is a 1 minute nasal airflow signal which begins 90 seconds before apnea occurs. The beginnings of apnea segments, automatically marked by apnea detection algorithm, are used to extract the pre-apnea interval. By this way, pre-apnea and apnea segments are disaggregated from each other by 30 seconds margin to minimize the pre-effects of apnea. Similarly, 1 minute signal intervals are extracted starting 90 seconds prior to healthy respiration, representing regular breathing. All of the regular breathing intervals, having shallow breathing for at most 2 seconds, are detected and sufficient number of these intervals are chosen randomly to be processed. By this way, it is intended to predict a possible apnea episode and alert before patient faces OSA, by having significant amount of time for processing. Time-series features of both classes are calculated and given as input features to learning algorithms. 131 pre-apnea intervals are extracted for each of 131 marked apnea segments. In order to decrease possible bias in the classification process, 120 (close to 131) regular breathing intervals are selected randomly, having a total of 251 input signals. Pre-apnea and regular breathing intervals are obtained from 6 patients to eliminate any personal effects among individuals.

3.1. Feature Selection

In this study only time-series features are considered and used. Most popular features like mean, variance, minimum, maximum, median values of signals are used not only for the original signal but also implemented for the power and derivative of the signal. In addition to these 15 features; minimum, maximum, average inspiration/expiration amplitudes and durations of nasal airflow signal are also extracted for each of 251 signal intervals (pre-apnea and regular breathing). As a result, a total of 39 features are extracted to be used in classification.

3.2. Support Vector Machines (SVM)

In this study, first implemented classification algorithm is Support Vector Machines (SVM). SVM is a popular and widely used classification algorithm in biomedical studies [11]. SVM basically seperates two different classes of inputs each represented in multi dimensional feature space by generating a boundary function. For the representation of different classes, regular breathing episodes are labelled as '0', as not an apnea episode, and pre-apnea episodes are labelled as '1', as an apnea episode. SVM algorithm is implemented by using the machine learning tool-box of MATLAB software.

3.3. k-Nearest Neighbors (kNN)

For the 2nd classification algorithm k-Nearest Neighbors (kNN) is used. kNN simply measures the distance of an input to all of the training dataset points in multi dimensional feature space. Then, the test input will be labelled as the class of the closest training point to itself. This approach is known as kNN-1. In another versions of kNN, defined as kNN-t (such as kNN-3, kNN-5 etc.), the closest t points to the test sample are detected and then a voting mechanism decides the class of the input. Consider a test sample that we would like to predict whether it is a pre-apnea segment. If two out of three closest points to the test sample belong to pre-apnea class, then this test input will be assigned to pre-apnea class as a result of 2-1 voting.

3.4. Linear Regression Approach for Classification (LR)

Linear Regression (LR) basically is a regression algorithm where the output is a real number, not a class. In this study, we approach this classification problem as regression. Similar to SVM learning, in training set pre-apnea episodes are labelled as '1' and regular breathing will be '0'. LR output will be a real number not only 0 or 1. In order to assign the test sample to a class, we should decide a level, call it l. The output higher than l value will be labelled as '1', and if the output is lower than l, then it will be assigned to class '0'.

3.5. Feature Reduction and Performance Measures

Selected 39 features are used for classification of a signal interval as pre-apnea or not. However, it is possible to decrease the dimension of the feature space and improve the classification accuracy by eliminating some of the irrelevant features. For this purpose, Randomly Select and Compute (RANSAC) algorithm is used. First, the accuracy of the classification is calculated by using all 39 features. Results are shown in Results & Discussion section. Then, 35 features out of 39 are randomly selected and accuracy is calculated using these 35 features. RANSAC algorithm is implemented for 1000 repetitions and features are noted having highest accuracy in classification. Similar feature reduction algorithm is computed for 30, 25, 20 features. Results suggest that randomly selected 30 features have the best performance in prediction of OSA episodes and further reduction of features also diminishes the performance. Feature reduction and classification algorithm is as follows:

choose number of features (Repeat for n=35,30,25 and 20)
❖ *randomly select n features (Repeat for 1000 times)*
✓ *train and classify by selected learning algorithm*
✓ *record features giving highest accuracy in selected learning algorithm*
✓ *5-fold cross validation for accuracy, sensitivity, specificity*

Performance measures are calculated using 5-fold cross validation. The whole data set is randomly divided into two categories: training set, test set. Training set contains 80% of the data set, and learning algorithm is trained by these input instances. The remaining 20% is used for classification and measuring the accuracy, sensitivity and specificity of the proposed approach. The training process repeated for 5 times and in each repetition, a new 80% training set is generated. After training and classifying the signals 5 times, the mean values are noted as final performance measures.

4. Results & Discussion

The performance of the proposed prediction algorithm is measured in terms of accuracy, sensitivity, specificity, accuracy for SVM, kNN & LR. Accuracy is the correct prediction of a pre-apnea interval (labelled '1') and regular breathing interval (labelled '0'). Sensitivity is the correct classification of pre-apnea intervals (labelled '1') and similarly specificity is the correct classification of regular breathing intervals (labelled '0'). Primarily, extracted 39 features are used for learning and classification process. 59% of accuracy is recorded which is below accepted levels. Followingly, feature reduction approach is used and performance is measured for using different features. Table 1 shows the performance measures of prediction methods for the given parameters.

Table 1: Performance measures of the proposed prediction methods.

Number of Features	SVM			kNN-3			kNN-5		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
39	0.591	0.880	0.514	0.602	0.589	0.620	0.749	0.763	0.740
35	0.839	0.894	0.798	0.635	0.656	0.624	0.785	0.776	0.792
30	**0.876**	**0.913**	**0.772**	**0.711**	**0.724**	**0.687**	**0.801**	**0.795**	**0.806**
25	0.853	0.920	0.724	0.685	0.707	0.662	0.784	0.804	0.743
20	0.812	0.905	0.746	0.685	0.716	0.632	0.745	0.779	0.708

Number of Features	LR (l=0.5)			LR (l=0.6)			LR (l=0.2)		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
39	0.528	0.726	0.504	0.496	0.724	0.481	0.733	0.786	0.700
35	**0.587**	**0.798**	**0.495**	**0.555**	**0.826**	**0.502**	0.745	0.808	0.692
30	0.533	0.777	0.472	0.511	0.832	0.461	**0.783**	**0.868**	**0.712**
25	0.505	0.688	0.440	0.504	0.776	0.473	0.716	0.794	0.638
20	0.486	0.643	0.421	0.486	0.709	0.443	0.724	0.760	0.653

As Table 1 states, all of the performance parameters show improvement after a few features are reduced. Elimination of unrelated features to the apnea prediction, increases prediction accuracy. However, further feature reduction has unfavorable effect on accuracy and the optimal results are obtained for set containing 30 features. Eliminated features are analyzed, and it is noticed that following features are always discarded in reduction process:

- ❖ maximum time between peak of inspiration and average value
- ❖ maximum time between peak of expiration and average value
- ❖ minimum time between average value and peak of expiration
- ❖ minimum time between average value and peak of inspiration
- ❖ ratio of time that respiratory signal is above the average value to the time respiratory signal is below the average value

We can comment that some of the features representing minimum or maximum time recorded for inspiration or expiration are generally eliminated in feature reduction. Thus, it can be concluded that extreme (minimum or maximum time) inspiration & expiration durations are not significant to predict OSA. Features providing information on general respiratory structure, like mean values inspiration/expiration or magnitude of nasal airflow signal, are more crucial and needed to be analysed especially.

SVM has the best prediction performance in accuracy, sensitivity, overall and kNN has the highest specificity rate. kNN-1 & kNN-7 are also tested for the selected features, however accuracy is lower than 0.5 and results are not added in the table since it does not show any significant meaning.

5. Conclusion

In an overnight sleep, OSA patients may face apnea more than 300 times. In addition to decrease in sleep quality and inability to maintain resting during sleep, OSA can be associated with other disorders like hypertension, stroke. It is an underestimated critical disorder, which BPAP or CPAP devices are used to overcome apnea. Having a sleep using these devices, discomforts the patients. In this study, an early warning algorithm to predict OSA episodes is proposed. In this manner, it may be possible to warn or wake the patient up before his/her respiration fails. The accuracy of apnea prediction is over 87% with a sensitivity of 91% when SVM is used. kNN approach has lowest accuracy but in return of highest specificity slightly above 0.8. For LR as a classification algorithm method, proper classification boundary selection (l value) significantly effects the general accuracy however, still it did not present any satisfactory results in any of the performance measures. 77% of specificity value of SVM suggests that there may be some cases, the algorithm labels a nasal airflow period as pre-apnea where actually it is not. Total computational time of the algorithm for a single 1 minute nasal airflow signal is less than 0.3 seconds. Thus, it may be useful and comfortable approach that a device, records the nasal airflow signal of the patient throughout the night and runs the proposed algorithm in real-time. By this way we may be able to warn or wake the patient up before his/her respiration fails. In future studies, specificity ratio will be tried to be improved. Development of a hybrid learning tool, combining the high accuracy, sensitivity performance of SVM and high specificity of kNN, will significantly improve the prediction of OSA segments.

References
[1] F. Roche, V. Pichot, E. Sforza, I. Court-Fortune, D. Duverney, F. Costes, M. Garet, and J-C. Barthe Iemy, "Predicting sleep apnea syndrome from heart period: a time-frequency wavelet analysis," *European Respiratory Journal.*
[2] P. E. Peppard, T. Young, J. H. Barnet, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults," *Am. J. Epidemiol.,* vol. 177, no. 9, pp. 1006-1014, 2013.
[3] M. R. Mannarino, F. Di Filippo, and M. Pirro, "Obstructive sleep apnea syndrome," *Eur. J. Intern. Med.,* vol. 23, no. 7, pp. 586-593, 2012.
[4] C. A. Kushida, M. R. Littner, T. Morgenthaler, C. A. Alessi, D. Bailey, J. Jr. Coleman, L. Friedman, M. Hirshkowitz, S. Kapen, M. Kramer, T. Lee-Chiong, D. L. Loube, J. Owens, J. P. Pancer, and M. Wise, "Practice parameters for the indications for polysomnography and related procedures: an update for 2005," *Sleep,* vol. 28, no. 4, pp. 499-521, 2005.
[5] Sleep Medicine Centers of Western New York. (2016, May 12) Available: http://www.sleepmedicinecenters.com/SleepDisorders/ObstructiveSleepAnea

[6]   R. Nowak, T. Corbridge, and B. Brenner, "Noninvasive Ventilation," in *Proceedings of the American Thoracic Society,* vol. 6, no. 4, pp. 367-370, 2009.

[7]   L. Chen, X. Zhang, and C. Song, "An Automatic Screening Approach for Obstructive Sleep Apnea Diagnosis Based on Single-Lead Electrocardiogram," *IEEE Trans. Autom. Sci. Eng.,* vol. 12, no. 1, pp. 106-115, 2015.

[8]   T. C. Huang, H. Y. Chen, and W. C. Fang, "Real-Time Obstructive Sleep Apnea Detection Based on ECG Derived Respiration Signal," *2012 IEEE International Symposium on Circuits and Systems (ISCAS),* 2012, pp. 341-344, 20-23.

[9]   J. Han, H. B. Shin, D. U. Jeong, and K. S. Park, "Detection of apneic events from single channel nasal airflow using 2nd derivative method," *Comput. Method. And Prog. Biomed.,* vol. 91, pp. 199-207, 2008.

[10]   S. I. Rathnayake, I. A. Wood, U. R. Abeyratne, and C. Hukins, "Nonlinear features for single-channel diagnosis of sleep-disordered breathing diseases," *IEEE Trans. Biomed. Eng.,* vol. 57, no. 8, pp. 1973-1981, 2010.

[11]   M. M. Nano, X. Long, J. Werth, R. M. Aarts, and R. Heusdens, "Sleep Apnea Detection Using Time-Delayed Heart Rate Variability," *Engineering in Medicine and Biology Society (EMBC), 37th Annual International Conference of the IEEE,* 2015, pp. 7679-7682.