

Sleep apnea classification based on respiration signals by using ensemble methods

Cafer Avci* and Ahmet Akbaş

Department of Computer Engineering, Yalova University, 77200, Yalova, Turkey

Abstract. In this study, an efficient and robust method classifying the minute based occurrence of sleep apnea is aimed. Three respiration signals obtained from abdominal, chest and nasal way extracted from polysomnography recordings. Wavelet transform based on feature extraction methods are applied on the 1 minute length respiration signals. Dimension reduction process is facilitated by using principal component analysis. The features obtained from 8 recordings are used for the classification sleep apnea by using three ensemble classifiers. According to the results, the classification accuracies have been obtained between 92.07-98.43%, 92.75-98.68% and 92.42-98.61% by using three different ensemble classifier based on abdominal, chest and nasal based analysis, respectively for AdaBoost, Random Forest and Random Subspace. However the best result is obtained analyzing nasal based respiratory signal by using Random Forest method. In this case accuracy is 98.68%.

Keywords: Sleep apnea, respiratory signals, wavelet, ensemble, boosting, bagging

1. Introduction

The sleep apnea syndrome (SAS) is asleep-related breathing disorder described as pauses in breathing or instances of shallow breathing during sleep suffered by people. SAS is occurred 2% in women and %4 in men whose age can range from 30 to 60 years [1]. Sleep apnea defined as the disruption of breathing for more than 10 s in adults [2]. Daytime sleepiness, impaired concentration, depression, memory loss, tiredness and early morning headaches can be observed in SAS patients [3-6]. Besides, untreated and undiagnosed sleep apnea can relate to heart disease like hypertension, myocardial infarction and stroke [7-9]. Sleep disorders have become important factors that adversely affect individuals' quality of life and society as a whole [10]. Therefore, treating the sleep-apnea is an important subject that may lead to certain diagnoses and better lifespan according to treatment.

Sleep apnea is basically divided into two categories; obstructive sleep apnea (OSA) and central sleep apnea (CSA). Respiratory drive is absent or inhibited in CSA. Upper airway collapses in responsible for disrupted respiration in OSA [11]. Some SAS subjects have both of them and in this case there is a pause of the respiratory followed by obstruction of ventilation in a relatively shorter

* Address for correspondence: Cafer Avci, Department of Computer Engineering, Yalova University, 77200, Yalova, Turkey. Tel.: +90 226 815 5342; Fax: +90 226 815 5401; E-mail: cafer.avci@yalova.edu.tr.

term, the event is called mixed sleep apnea (MSA) [12]. To diagnose of SAS, the most reliable method is the analysis of the patient's polysomnography (PSG). PSG consists of several multiple vital signals; body oxygen saturation (SaO_2), electrooculogram (EOG), electrocardiogram (ECG), electroencephalogram (EEG), electromyogram (EMG), respiration signals (nasal airflow, thoracic abdominal movements), body position. According to the accepted clinical criteria, sleep apnea can primarily be detected by the analysis of PSG [13]. Manually scoring PSG is a tiresome and time-consuming work for the sleep expert. The subjectivity and repetitivity of the work expose low inter-scoring agreement and accuracies. Computerized decision-making methods can automate these routine processes.

Various features extracted from respiration signals suitable for detection of sleep apnea by using non-linear [14], time-frequency [15], time [16] and frequency domain [17] algorithms have already been studied. These algorithms can detect episodes of apnea at 80-95% precision. Tagluk, et al. studied a classification of sleep apnea based on abdominal respiration signal by using wavelet transforms and neural networks [18]. These algorithms have 85.62% mean accuracy in classifying sleep apnea. Koley and Dey have been developed automatic detection method using ensemble binary support vector machine classifiers [19]. 36 features have been used to detect sleep apnea. In this method, accuracies of algorithm change between 86% and 94.1.

It is aimed to develop sleep apnea detection method and evaluate performance of method by using three different respiration signals. Features are obtained by using wavelet decomposition of signals and they are classified by using ensemble methods for the real time sleep apnea detection. Authors have published studies in the field of detection of sleep apnea [20-22].

New results are presented in this study for detection of sleep apnea using ensemble learning methods. In Part 2, materials and methods are introduced including data explanation in detail, wavelet based feature extraction and ensemble methods. Performance metrics and experimental results are described in Part 3 and the study is concluded in Part 4.

2. Materials & methods

2.1. Data set

In this study, the data set used for the evaluation is Apnea-ECG Database, which is available on the PhysioNet website [23]. Database has 70 ECG recordings digitized 100 Hz from normal and OSA patients which duration varying between 401 and 578 minutes. Data were gathered in Philips University in Marburg, Germany having sleep apnea annotations was done by sleep experts according to standard criteria simultaneously recorded respiration signals [24].

Only 8 of the recordings (7 M and 1F, 43.3 ± 8.3 years) have respiration signals (nasal, abdominal and chest) in the database. They are used to detect apnea episodes. Each minute of the recordings was labeled with 'A' or 'N', indicating the apneic or non-apneic (normal) respectively during that minute. The apnea/hypopnea standards (AHI) are used to classify an epoch as apneic or normal by counting the number of apneic episodes over the given period of time and averaging these counts on a per hour basis [25]. The recordings are classified into three class; apnea class (class A, SAS > 100 minutes), borderline class (class B, SAS > 5 and SAS < 100 minutes) and normal class (class C, SAS < 5 minutes). The recordings used in this study have 3 Class C patients, 1 Class B and 4 Class A. A 3 minutes length of abdominal, chest and nasal based respiratory signals with minute based annotations ('N' normal, 'A' apneic) are shown in Figure 1.

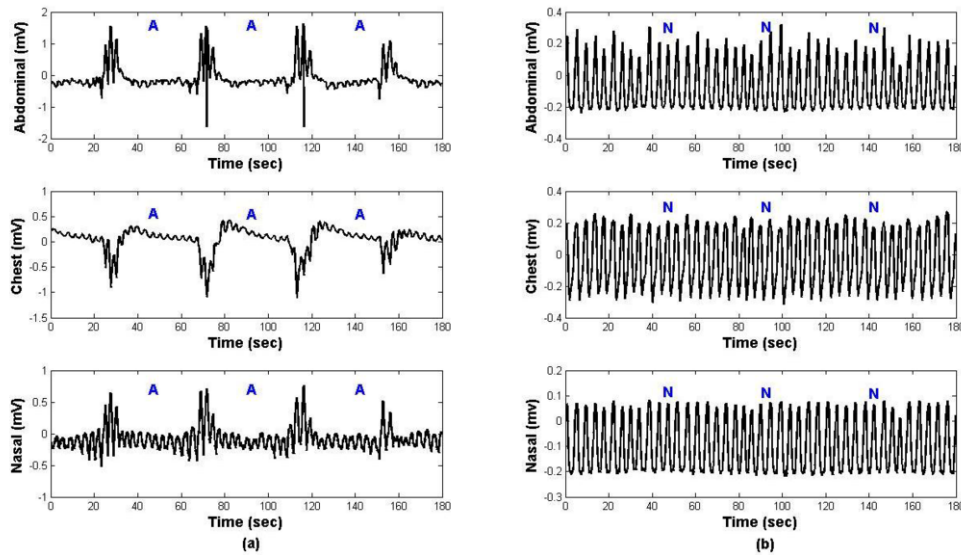


Fig. 1. 3 min. length respiratory signals (apneic: a, non-apneic: b).

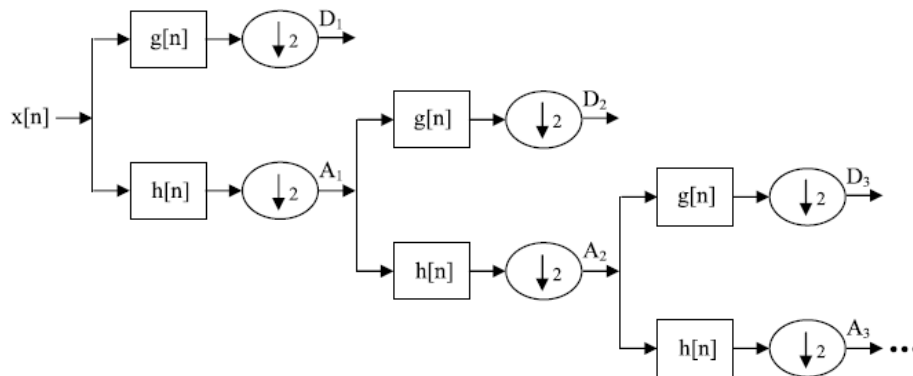


Fig. 2. Decomposition of discrete wavelet transform; $g[n]$ is the high-pass filter (Detail), $h[n]$ is the low-pass filter (Approximation).

2.2. Feature extraction by wavelet analysis

The Wavelet transform operates by decomposing a signal into a number of time shifted and scaled versions of a selected mother wavelet. Wavelet transform is capable of providing the time and frequency information simultaneously, hence giving a time-frequency representation of the signal. They are suitable for the analysis of non-stationary signals giving good results both in frequency and time domain. The basic implementation of wavelet transform of a signal $x[n]$ is schematically shown in Figure 2 [26].

There are many different functions as wavelets having different characteristics that are more or less suitable depending on the application. Daubechies wavelet family with a length of the filter equal to 3 was chosen after performing some other experiments with other wavelet families. For the classification process; minimum, maximum, variance, average, energy, mode, entropy and skewness have been extracted. All features have been calculated after decomposing the 1-minute length of detail

components for all 11 levels obtained from dataset consisting of 8 recordings. Totally, 88 features have been extracted for 3325 minutes respiration signals obtained from all recordings.

For the dimension reduction, correlation based feature subset selection (CfsSubsetEval) is used as feature selection algorithm. It evaluates subset measures of features by using following hypothesis: good feature subsets include highly related features with classification more than irrelevant one [27]. 12 features are used for the classification process after applying CfsSubsetEval.

2.3. Classification

Ensemble learning [28] technique is derived from the principles of regular machine learning topic. Ensemble methods use multiple learning algorithms to obtain better results than conventional learning algorithms. Ensemble learning is a technique for combining a lot of base learner in an attempt to produce better result. Basic machine learning classifiers such as Bayes, Decision Trees, etc. is called base learners or weak classifier in ensemble methods. Three ensemble methods were implemented in this study. Majority voting is used to select final decision of base learners. Methods are briefly explained as follows.

2.3.1. Random forest (RF)

Random Forest is an ensemble of decision tree classifiers and a variation of Bagging. Random Forest has several Decision Trees as base learners, which are learning from bootstrapping samples of the training data set and uses random feature selection in the tree induction process. The main strategy of the bagging is changing the dataset by resampling to train weak learners. Changed samples are carried out during voting process of weighted parameters. Each tree classifier generates a vote for the final decision of the forest and the class of the input. The Random Forest then chooses the classification having the most votes over all the trees in the forest. Random Forest has more advantages than the other learners, such as high accuracy, robustness to noise and outliers, implementation simplicity, easy to cope with over fitting. For the classification of sleep apnea, 100 Decision Tree with Bagging resampling method is implemented to detect minute is either apneic or non-apneic.

2.3.2. Adaptive boosting (AdaBoost)

AdaBoost is extremely important ensemble algorithm by selecting weak learners. AdaBoost finds a combination of weak classifier with updating weight through repetitive process without changing the original training data set. Algorithm is given below:

Input:

$(x_1, y_1), \dots, (x_m, y_m)$: sample series

$y_i \in Y = \{1, \dots, k\}$: class data

L : learning algorithm

T: iteration number

Start: for every $D_1(i) = 1/m$.

Do for these values: $t=1, 2, \dots, T$:

* Call for L by carrying out distributions with obtained D_t values.

* Hypothesis $h_t : X \rightarrow Y$

* Calculation of error for the hypothesis

$$h_t : \epsilon_T = \sum_{i:h_t(x_i) \neq (y_i)} D_t(i)$$

If $\epsilon_T > 0.5$ then update T to $T = t - 1$ and exit loop.

- * Update β_t value $\beta_t = \epsilon_T / (1 - \epsilon_T)$.
- * Update distribution value D_t :

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} x \begin{cases} \beta_t, h_t(x_i) = y_i \\ 1, \text{other process} \end{cases}$$

Z_t : Normalization constant
 Output: Final hypothesis

$$h_{fin}(x) = \underset{y \in Y}{\operatorname{arg\,max}} \sum_{t:h_t(x)=y} \log \frac{1}{\beta_t}$$

2.3.3. Random subspace (RSS)

Random Subspace is a generalization of the Random Forest algorithm. It consists of Decision Trees combination whereas Random Subspace consists of any classifier. In this study, Decision Trees are used in Random Subspace as base learners. An identical number of base learners with other ensemble models are used.

3. Results

3.1. Performance metrics

Comparisons have been made by using evaluation methods that are accepted in literature to measure reliability of results, in this study. All notations that used for the performance evaluations are showed as confusion matrix in Table 1.

- * TP: Number of true positives.
- * TN: Number of true negatives.
- * FP: Number of false positives.
- * FN: Number of false negatives.

3.1.1. Accuracy

Accuracy refers to the closeness of a measured value to a standard or known value. Accuracy is calculated with following equations;

Table 1
 Confusion Matrix

	Prediction		
Actual	TP	FP	Precision
	FN	TN	
	Recall		Accuracy

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3.1.2. F-measure

Recall and precision are accepted as performance metrics that resolve false positive and false negative errors. Recall and precision equations are given below.

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

F-measure also known as F1 Score can be derived from Recall and Precision measures with equation given below. Weighted average of the Recall and Precision defines the F-measure. F-Measure gives best score at 1 and worst score at 0.

$$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

3.1.3. Kappa

Kappa is a well-known method for analyzing ensemble methods. Kappa is traditionally used to quantify the degree of agreement between two raters on a nominal scale. Difference between zero shows fineness of result. Kappa is calculated with following equation;

- * K : Kappa result.
- * $Pr(a)$: The proportion of units where there is agreement.
- * $Pr(e)$: The proportion of units which would be expected to agree by chance.

$$Kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

3.2. Classification results

Totally 3325 1-minute based feature vectors are obtained from annotated data set. For the training and testing adjustment, 10-fold cross validation method is used. The performance metrics according to the obtained 1-minute based classification results for abdominal, chest and nasal based respiratory signals by using AdaBoost, Random Forest and Random Subspace are given in Table 2, respectively.

According to the results, the classification accuracies have been obtained between 92.07-98.43%, 92.75-98.68% and 92.42-98.61% by using three different ensemble classifiers based on abdominal,

Table 2

Classification results of sleep apnea based on 1-minute analysis of respiration signals

Classifier	Classification Accuracies (%)			F-Measure			Kappa		
	Abdominal	Chest	Nasal	Abdominal	Chest	Nasal	Abdominal	Chest	Nasal
Ada Boost	92.07	95.50	98.43	0.911	0.955	0.979	0.916	0.938	0.970
Random Forest	92.75	95.87	98.68	0.928	0.959	0.981	0.925	0.936	0.972
Random Subspace	92.42	95.45	98.61	0.914	0.945	0.972	0.916	0.937	0.971

chest and nasal based analysis, respectively for AdaBoost, Random Forest and Random Subspace. However, the best result is obtained analyzing nasal based respiratory signal by using Random Forest method. In this case, accuracy, F-Measure and Kappa are obtained 98.68%, 0.981 and 0.972, respectively.

4. Discussion and conclusions

According to the results given in Table 2, all classification results show that developed method can be used for obstructive sleep apnea detection. According to the results, success ranking of SAS detection based on three different respiration signals is nasal, chest and abdomen, respectively. However, present methods because of their robustness and stability as using only wavelet transforms give methods the impact to the rank of the method. Polysomnography systems or embedded systems including presented methods for automatic OSA detection can improve the medical service efficiency. We will focus on to develop a sleep apnea detection method based on only ECG signal by using larger data set, in our future research.

References

- [1] S. Ryan and W.T. McNicholas, Obstructive Sleep Apnea, in: International Encyclopedia of Public Health, Harald Kristian (Kris) Heggenhougen, ed., Academic Press, Oxford, 2008, pp. 621-627.
- [2] Teresa Paiva and Hrayr Attarian, Obstructive sleep apnea and other sleep-related syndromes, in: Handbook of Clinical Neurology, José Biller and José M. Ferro, eds., Elsevier, [Netherlands](#), 2014, pp. 251–271.
- [3] Graziela De Luca Canto, Camila Pachêco-Pereira, Secil Aydınoç, Paul W. Major, Carlos Flores-Mir and David Gozal, Biomarkers associated with obstructive sleep apnea and morbidities: A scoping review, *Sleep Medicine* **16** (2015), 347–357.
- [4] Lenise Jihe Kim, Denis Martinez, Cintia Zappe Fiori, Diego Baronio, Nélson Alexandre Kretzmann and Helena Maria Tannhauser Barros, Hypomyelination, memory impairment, and blood–brain barrier permeability in a model of sleep apnea, *Brain Research* **1597** (2015), 28–36.
- [5] J.C. Rains and J.S. Poceta, Headache and sleep, in: Encyclopedia of Sleep, Clete A. Kushida, ed., Academic Press, Waltham, 2013, pp. 329-333.
- [6] Melanie Harris, Nick Glozier, Rajeev Ratnavadivel and Ronald R. Grunstein, Obstructive sleep apnea and depression, *Sleep Medicine Reviews* **13** (2009), 437–444.
- [7] D.E. Wardly, Intracranial hypertension associated with obstructive sleep apnea: A discussion of potential etiologic factors, *Medical Hypotheses* **83** (2014), 792–797.
- [8] Filip M. Szymanski, Grzegorz Karpinski, Anna E. Platek and Krzysztof J. Filipiak, Acute myocardial infarction type 2 secondary to the obstructive sleep apnea, *Coret Vasa* **55** (2013), e449–e452.
- [9] Owen D. Lyons and Clodagh M. Ryan, Sleep apnea and stroke, *Canadian Journal of Cardiology* **31** (2015), 918-927.
- [10] R.N. Aurora, Health-related quality of life in obstructive sleep apnea syndrome, in: Encyclopedia of Sleep, Clete A. Kushida, ed., Academic Press, Waltham, 2013, pp. 424–427.
- [11] V.K. Somers, D.P. White and R. Amin, Sleep apnea and cardiovascular disease: An American heart association/American college of cardiology foundation scientific statement from the American heart association council for high blood pressure research professional education committee, council on clinical cardiology, stroke council, and council on cardiovascular nursing, in collaboration with the national heart, lung, and blood institute national center on sleep disorders research (national institutes of health), *Circulation* **119** (2008), 118–1080.
- [12] P. de Chazal, T. Penzel and C. Heneghan, Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram, *Physiological Measurement* **25** (2004), 967–983.
- [13] P. Varady, S. Bongar and Z. Benyo, Detection of airway obstructions and sleep apnea by analyzing the phase relation of respiration movement signals, *IEEE Transactions on Instrumentation and Measurement* **52** (2003), 2–6.
- [14] S.I. Rathnayake, I.A. Wood, U.R. Abeyratne and C. Hukins, Nonlinear features for single-channel diagnosis of sleep-disordered breathing diseases, *IEEE Transactions on Biomedical Engineering* **57** (2010), 1973–1981.

- [15] P. Caseiro, R. Fonseca-Pinto and A. Andrade, Screening of obstructive sleep apnea using Hilbert–Huang decomposition of oronasal airway pressure recordings, *Medical Engineering and Physics* **32** (2010), 561–568.
- [16] B.L. Koley and D. Dey, Real-time adaptive apnea and hypopnea event detection methodology for portable sleep apnea monitoring devices, *IEEE Transactions on Biomedical Engineering* **60** (2013), 3354–3363.
- [17] H. Nakano, T. Tanigawa, T. Furukawa and S. Nishima, Automatic detection of sleep-disordered breathing from a single-channel airflow record, *European Respiratory Journal* **29** (2007), 728–736.
- [18] M. Tagluk and N. Sezgin, Classification of sleep apnea through sub-band energy of abdominal effort signal using wavelets + neural networks, *Journal of Medical Systems* **34** (2010), 1111–1119.
- [19] Bijoy Laxmi Koley and Debangshu Dey, Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary SVM classifiers, *Measurement* **46** (2013), 2082–2092.
- [20] C. Avcı and A. Akbaş, Comparison of the ANN based classification accuracy for real time sleep apnea detection methods, *Proceedings of the IASTED International Conference Biomedical Engineering (BioMed 2012)*, Innsbruck-Austria, 2012, pp. 56-60.
- [21] C. Avcı, I. Delibasoglu and A. Akbas, Sleep apnea detection using wavelet analysis of ECG derived respiratory signal, *2012 International Conference on Biomedical Engineering (ICoBE)*, Penang, Malaysia, February 27-28, 2012, pp. 272–275.
- [22] C. Avcı and G. Bilgin, Sleep apnea detection using adaptive neuro fuzzy inference system, *Engineering* **5** (2013), 259–263.
- [23] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C.H. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng and H.E. Stanley, Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals, *Circulation* **101** (2000), e215–e220.
- [24] T. Penzel, G.B. Moody, R.G. Mark, A.L. Goldberger and J.H. Peter, The apnea-ECG database, *Computers in Cardiology* **27** (2000), 255–258.
- [25] W.R. Ruchland, P.D. Rochford, F.J. O'Donoghue, R.J. Pierce, P. Singh and A.T. Thornton, The new AASM criteria for scoring hypopneas: Impact on the apnea hypopnea index, *PubMed Central Journals, Sleep* **32** (2009), 150–157.
- [26] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, *IEEE Transaction on Inform Theory* **36** (1990), 961–1005.
- [27] M.A. Hall, Correlation-based feature subset selection for machine learning, Ph.D. Dissertation, University of Waikato, 1998.
- [28] G. Dietterich Thomas, Ensemble methods in machine learning, multiple classifier systems, *Lecture Notes in Computer Science* **1857** (2000), 1–15.