

k-NN-based classification of sleep apnea types using ECG

Oğuzhan TİMÜŞ¹, Emine DOĞRU BOLAT^{2,*}

¹Inventory Control Center, Turkish Navy, Gölcük, Kocaeli, Turkey

²Department of Biomedical Engineering, Faculty of Technology, Kocaeli University, Kocaeli, Turkey

Received: 10.11.2015

Accepted/Published Online: 26.11.2016

Final Version: 30.07.2017

Abstract: Obstructive sleep apnea syndrome (OSAS) is a common sleep disorder that yields cardiovascular diseases, excessive daytime sleepiness, and poor quality of life if not treated. Classification of OSAS from electrocardiograms (ECGs) is a noninvasive method and much more affordable than traditional methods. This study proposes a pattern recognition system for automated apnea diagnosis based on heart rate variability (HRV) and ECG-derived respiratory signals. The k-nearest neighbor (k-NN) classifier has been used to develop the models for classifying the sleep apnea types. For comparison purposes, classification models based on multilayer perceptron, support vector machines, and C4.5 decision tree (C4.5 DT) have also been developed. The first database used for training contains 12 and the second used for testing contains 35 whole-night polysomnography recordings from real subjects. Wrapper-based feature selection, optimal parameter calculation, and 10-fold cross-validation were applied to the training dataset. The performance of the classifiers was evaluated by accuracy, sensitivity, and specificity metrics. The k-NN classifier yields higher classification accuracy, sensitivity, and specificity by successfully separating 100% of apnea recordings from normal recordings, and it also achieves a classification rate of 97% accuracy, 89% sensitivity, and 100% specificity of the subjects in the test database. Median, mean, absolute deviation, and interquartile range values of HRV were the most descriptive parameters. These results indicate significant potential for achieving basic estimates for OSAS patients.

Key words: Electrocardiogram, heart rate variability, sleep apnea, k-nearest neighbor, classification, wrapper methods

1. Introduction

Sleep is of vital importance, accounting for almost one-third of each human's life. Lack of sleep or presence of sleep disorders prevents individuals' efficient use of sleep function, resulting in health problems related to the immune system, daily physical activity, stress, and concentration. Furthermore, sleep disorders may also cause various medical problems, including obesity, chronic heart failure, and challenges related to menopause, and they may increase cardiovascular mortality in conjunction with these problems [1].

Sleep apnea is defined as a decrease in or a complete interruption of breathing for at least 10 s, and it is the most frequently encountered problem among sleep disorders. It is categorized into three forms: obstructive, central, and mixed [2,3]. This interruption of airflow causes the body not to produce the necessary essential hormones and prevents bodily regeneration, making the affected individual's life unhealthy, unrestful, and unbearable. Sometimes, it may even result in death during sleep.

The apnea-hypopnea index (AHI) or the respiratory disturbance index (RDI) is employed as an index for evaluating sleep quality. AHI shows the total number of apneas and hypopneas per hour. Sleep apneas are

*Correspondence: ebolat@gmail.com

diagnosed utilizing the AHI value in the following way: AHI values < 5 are designated as normal or control, AHI values between 5 and 15 are grouped as borderline or mild apnea, while AHI values > 15 are categorized as sleep apnea [3].

Polysomnography is a standard technique used for the recognition of sleep apnea by recording a large number of physiological parameters such as electrocardiography (ECG), electro-encephalography (EEG), electrooculography (EOG), electromyography (EMG), blood oxygen saturation (SpO_2), and respiratory airflow. While the patient is sleeping in a sleep unit, these physiological parameters are recorded. Then a specialist observes these recordings visually and performs the scoring process [4].

Although sleep apnea appears to be a respiratory activity, it inevitably has an influence on the cardiovascular system. Hence, the ECG supplies quite useful data in cases of apneic problems [3,5]. The ECG signal is easy to be processed for diagnosis of sleep apnea. Additionally, the diagnosis of sleep apnea from the ECG signal is more practical and less expensive, requiring only ECG measuring devices.

Heart rate variability (HRV) is based on the sequence of RR-intervals extracted from the ECG, and some deviations from normal patterns during sleep may be due to sleep apnea. Thus, researchers have been performing various studies based on RR-intervals of the ECG signal [3,6–10]. In addition to HRV, ECG-derived respiratory (EDR), which is a representative respiration signal, is also derived from ECG and is used for apnea detection [5,11–15].

The aim of this study is to propose a novel, fast, and reliable pattern recognition system for automated apnea diagnosis. The novelty and significance of this study is the application of only time-domain parameters of HRV and EDR obtained from single-lead ECG recordings sourced from two different apnea databases. This study is built on 17 features and the k-nearest neighbor (k-NN) classifier. For comparison purposes, classification models based on MLP, SVM, and C4.5 DT have also been developed. The wrapper-based feature selection technique was employed for obtaining the best possible success with the classifier to be used. The classifiers' optimization was assessed by the classification accuracy, sensitivity, and specificity. For training the classifiers, a 10-fold cross-validation method was utilized. The classifiers were tested, and the classification results for the subjects were compared.

2. Related works

In 1999, Roche et al. analyzed the discriminant power of time-domain HRV parameters belonging to 5-min segments and concluded that time-domain HRV analysis might be an appropriate and inexpensive tool for obstructive sleep apnea syndrome (OSAS) diagnosis [8]. In addition, in 2003, Roche et al. employed wavelet transform on time-frequency parameters of HRV for a group of 147 subjects and achieved 92.4% sensitivity and 90.1% specificity [9].

Yilmaz et al. conducted a study on sleep stage and apnea classification in 2010 [3]. They employed the interquartile range (IQR), the mean absolute deviation (MAD), and the median parameters derived from the HRV. Quadratic discriminant analysis (QDA), k-NN, and SVM were taken as classifiers, and a 94.5% accuracy rate was achieved from the QDA- and SVM-based classifications.

ECG-based obstructive sleep apnea detection was proposed by Chazal et al. in 2003 [11]. They utilized linear discriminant analysis (LDA) and QDA for the features extracted from the HRV and the EDR signals. The subject group consisted of 70 normal, borderline apnea, and apnea recordings. They obtained 100% success for separating apnea recordings from normal recordings and over 90% accuracy for minute-by-minute classification.

In a follow-up effort in 2004, they conducted a study on the detection of obstructive sleep apnea at different time scales including 15, 30, 45, 60, 75, and 90 s using HRV and EDR features, and they obtained a 91% accuracy rate using the LDA [12]. Isa et al. [10] analyzed two approaches and concluded that the classification outcomes employing features suggested by Yilmaz et al. [3] provide 3.59% higher classification accuracy and 7.5% higher area under the ROC curve (AUC) value than the 8 features suggested by Chazal et al. [12]. In 2007, Mendez et al. studied the power spectral density of the R peak areas and the HRV utilizing a bivariate autoregressive model. For distinguishing the apnea events from normal ones, k-NN was employed, and classification accuracy higher than 85% was achieved [13]. In a follow-up investigation in 2009, they utilized a bivariate time-varying autoregressive model to appraise beat-by-beat power spectral density for the QRS complex areas and the RR-intervals. Neural network (NN) and k-NN classifiers were used to classify apnea and nonapnea minutes. A sequential forward selection was utilized to choose the best feature subset in a wrapper setting. Both k-NN and NN achieved a classification accuracy of 88% [5]. In 2009, Khandoker et al. [14] conducted a study on the automatic diagnosis of OSAS using wavelet-based parameters derived from both EDR and HRV employing the SVM classifier. With the Cohen's κ value of 0.85, they achieved classification accuracy of 92.85%. In 2011, Yildiz et al. conducted a study on OSAS diagnosis using HRV and EDR parameters for 30 subjects [15]. They used wavelet transform (DWT) and fast-Fourier transform (FFT) methods for feature extraction, and least squares SVM (LS-SVM) for classification, and achieved 100% accuracy for separating normal subjects from subjects with OSAS.

The similarities and differences of this work relative to the remaining related studies in the literature are described below and detailed in the Discussion section.

In the present study, multiple databases were utilized for apnea detection from ECG signals, similar to [5,14,16]; however, the majority of the related work consists of recordings from a single database [3,10–13,15,17,18]. The related work is mostly based on features obtained from HRV only [3,8–10,17,18], but in the present study, features from HRV and EDR were employed together, since they provide better classification performance than using HRV features alone [11,14]. In the related work, HRV and EDR parameters were mostly analyzed in the frequency domain [5,13–16]. In the present study, the parameters were analyzed in the time domain, similar to [14]. The HRV parameters used in our work are the most commonly used cardio-respiratory features, as advised by the Task Force of ESC and the NASPE for time-domain analysis of HRV. The EDR signal was obtained by summing 50 ms of data on either side of each R-peak, similar to [11,15]. The time scale of 1 min for epochs was chosen in this work, as in [5,7,10–15].

3. Materials and methods

The ECG recordings utilized for training and testing were gathered from two databases. These recordings were scored in apneic events as central, obstructive, mixed, and hypopnea by a specialist according to the American Academy of Sleep Medicine (AASM) guidelines [19]. For identifying the existence or absence of apnea/hypopnea for each minute of the data, an interactive algorithm was employed. Each minute-long epoch was described as apneic if one apnea or hypopnea event happened; otherwise, it was described as normal. This process was performed for the total sleep period for all data in two databases.

Figure 1 illustrates the schematic diagram of the methodology utilized in this paper. First, the PSG recordings for each subject were imported to the MATLAB environment. For processing and detecting R-peaks, the ECG recordings were divided into 1-min-long epochs. Each recording was visually inspected, and 128 epochs containing high-amplitude artifacts mainly caused by body movements were discarded. For obtaining

the missed R-peaks and for excluding the false R-peaks, a correction algorithm was employed before the HRV and EDR calculations.

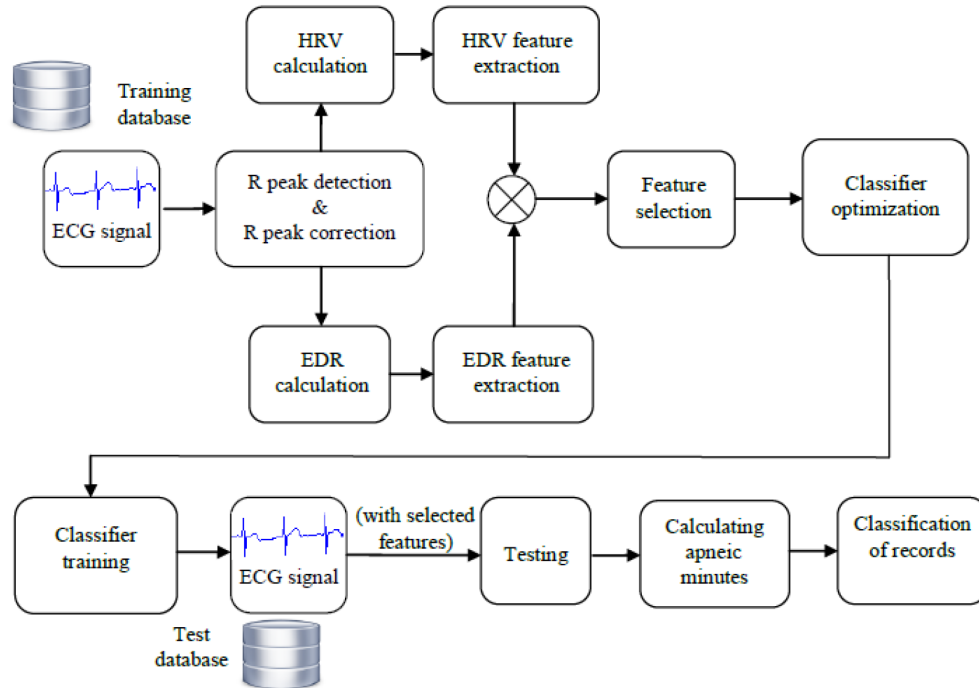


Figure 1. Schematic diagram of the methodology [32].

In the present study, time-domain features obtained from HRV and EDR data were utilized. To achieve the best possible success with the classifiers, the wrapper-based feature selection method was utilized, employing classifier accuracy as an evaluation measure. The optimal parameters for the classifiers were calculated before the training procedure. On the training dataset, a 10-fold cross-validation technique was employed in the training procedure. The trained classifiers were tested with reference to the test database. The recordings in the test dataset were classified into three classes as follows:

Class A (apnea): Recordings with $AHI > 15$. This class contains 15 subjects.

Class B (mild/borderline apnea): Recordings with $5 < AHI < 15$. This class contains 12 subjects.

Class C (normal): Recordings with $AHI < 5$. This class contains 8 subjects.

Eventually, the success of the classification used in this study was assessed with the metrics of accuracy, sensitivity, and specificity. The classification accuracy commonly used in pattern recognition is calculated as the ratio of the number of subjects correctly classified to the total number of subjects. Accuracy, sensitivity, and specificity are provided by Eqs. (1), (2), and (3), respectively [13]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$Specificity = \frac{TN}{FP + TN} \times 100\% \quad (3)$$

TP: The number of true positives, with TP defined as an input determined as a subject having apnea that was also recognized by the specialists.

TN: The number of true negatives, with TN defined as an input determined as a subject identified as healthy and who was recognized as healthy by the specialists as well.

FP: The number of false positives, with FP defined as an input determined as a subject identified to have apnea, but recognized as healthy by the specialists.

FN: The number of false negatives, with FN defined as an input determined as a subject identified as healthy, but recognized to have apnea by the specialists [15].

3.1. ECG recordings used in this study

Two different databases were used for the ECG recordings in this paper. The recordings taken from the first database were employed for training, while the recordings gathered from the second database were utilized for testing. The first database from which recordings were utilized for training is called the DREAMS Project, activated by the University of MONS - TCTS Laboratory and the Université Libre de Bruxelles - CHU de Charleroi Sleep Laboratory (<http://www.tcts.fpms.ac.be/~devuyt/Databases/DatabaseApnea/>); it is intended for automatic sleep signal analysis of adults [20]. The database used for training includes 12 whole-night PSG recordings taken from patients having sleep apnea trouble and scored by a specialist for apneic events such as central, obstructive, and mixed apnea and hypopnea. The selected data were obtained from a sleep laboratory in a Belgian hospital utilizing a digital 32-channel polygraph (Brainnet System of MEDATEC, Brussels, Belgium). The recordings contain EOG, EEG, EMG, ECG, the nasal airflow, and the abdominal and thoracic inductive plethysmography records. The sampling frequency was 200 Hz. The scoring process was applied using the AASM criteria. The age of the subjects varied between 41 and 79 years (53.93 ± 7.6), sleep duration varied between 7.5 and 9.5 h (8.63 ± 0.7), and AHI varied from 7.4 to 80.9 apnea/hypopnea per hour (46.52 ± 25.9).

The second database from which recordings were utilized for testing was obtained from the Kocaeli University Hospital Sleep Laboratory employing a digital 44-channel polygraph (Compumedics E-series).

The analyzed data were the data employed for the sleep laboratory's standard diagnosis and treatment process and scored by the specialist for obstructive, central, and mixed apnea as well as hypopnea events. The sampling frequency for ECG acquisition was 256 Hz, and the values of the band-pass filter cutoff frequency were selected as 0.3 Hz and 30 Hz, respectively. The PSG recordings contain 35 (21 men) subjects' whole-night recordings covering a range between 6.42 and 7.99 h (7.15 ± 0.47), the subject age between 27 and 78 years (47.85 ± 12.01), and AHI between 0.3 and 66.1 events per hour (19.29 ± 14.97). It is worth noting that the ECG recordings employed in this study were selected from subjects with no previously known heart disease and relatively clear signals preferred, to eliminate the miscalculation of heartbeats.

File reader code was employed to provide the PSG signals saved in standard European data format (EDF) and transfer to the MATLAB environment [21].

3.2. HRV and EDR calculation

Various studies for sleep apnea detection based on HRV have been performed due to the simplicity and efficiency of ECG. HRV supplies many parameters in both the frequency domain and the time domain, and deviations from normal patterns of HRV during sleep can be used to detect the presence of sleep apnea. HRV is the temporal variation between sequences of consecutive heartbeat intervals. RR intervals should be obtained to permit the calculation of HRV from the ECG signal. The RR interval is the period between two adjacent R

waves [22,23]. Many R-wave detection approaches can be found in the literature [14,15,24–27]. These techniques can be summarized as Teager energy operator (TEO), wavelet transform, SVM, the Pan–Tompkins algorithm, filtering in the frequency domain, and zero crossing rate [28]. Since TEO is the most commonly used approach utilized for numerous applications such as one-dimensional signals, images, and color image processing [29] and is convenient for studying a single component in both continuous and discrete domains, it was used in this study to obtain the RR intervals. In the continuous and discrete domains, TEO is expressed as in Eqs. (4) and (5), respectively [28,29]:

$$\psi_c [x(t)] = \dot{x}^2(t) - x(t) \times \ddot{x}(t) \quad (4)$$

$$\psi [x(n)] = x^2(n) - x(n-1) \times x(n+1) \quad (5)$$

Over the signal off-line, TEO is applied for the discrete domain, employing a specific sampling frequency value. TEO, spanning the three adjacent samples of the signal, indicates a very local feature of the signal [30]. The peak finder function was applied in each epoch to provide the R-wave peak values. This function provided the value and the location of the peak value in a range of 0.6 s of the sampling rate (120 samples for training dataset and 154 samples for test dataset). Figure 2a depicts the ECG signal before and Figure 2b illustrates the ECG signal after TEO was performed. Figure 2c shows the obtained R-wave peaks. After missed or false R peaks were determined employing a correction algorithm, the HRV value of each epoch was obtained from the intervals between detected consecutive R peaks. EDR is a representative respiration signal derived from ECG. The difference in thoracic electrical impedance and the movement of the electrode relative to the heart throughout the breathing cycle impacts the body surface ECG [11]. This impact can be seen in the amplitude of the baseline-corrected ECG signal. Prior to the derivation of EDR, baseline correction was applied to the ECG signal. Later, the QRS area, with a duration of 100 ms, was calculated by summing 50 ms of data (10 samples for 200 Hz sampling rate) of either side of each R peak. An example for the correlation of the EDR signal with the respiratory signal is shown in Figure 3, where Figure 3a is the respiration signal, Figure 3b is the EDR signal derived from ECG, and Figure 3c is the baseline-corrected ECG signal.

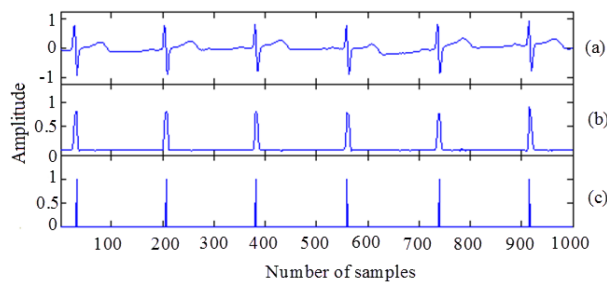


Figure 2. (a) ECG signal. (b) ECG signal after TEO. (c) Detected R-wave peaks [32].

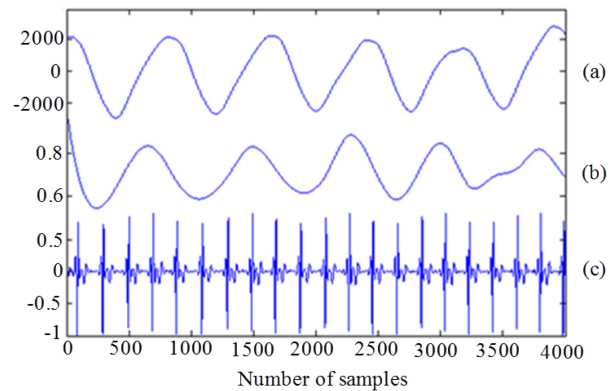


Figure 3. (a) Respiration signal. (b) EDR signal (c) Baseline-corrected ECG [32].

3.3. Feature extraction

The features in this study were extracted from the HRV and EDR data and analyzed in the time domain [7]. The features are listed as follows:

- inter-quartile range (IQR), the measure of statistical dispersion equal to the difference between the upper (75th) and lower (25th) quartiles
- mean absolute deviation (MAD), the mean value of absolute difference of the dataset from its mean
- median, mean, and standard deviation (SDNN) of RR interval
- root mean square of successive differences (RMSSD), the square root of the mean of the sum of the squares of the successive differences between adjacent RR intervals
- standard deviation of successive differences (SDSD), the standard deviation of the successive differences between adjacent RR intervals
- NN50 count (variant 1), the number of all pairs of adjacent RR intervals differing by more than 50 ms
- NN50 count (variant 2), the number of pairs of adjacent RR intervals differing by more than 50 ms in which the **first** interval is longer
- NN50 count (variant 3), the number of pairs of adjacent RR intervals differing by more than 50 ms in which the **second** interval is longer
- pNN50 (variant 1), pNN50 (variant 2), and pNN50 (variant 3), the proportion of each NN50 count divided by the total number of RR intervals
- \overline{HR} , the mean of the heart rate
- IQR, median, and MAD values of EDR.

The features, SDNN, RMSSD, SDSD, NN50 count variants, and pNN50 variants, were advised by the Task Force of the European Society of Cardiology (ESC) and the North American Society of Pacing and Electrophysiology (NASPE) for time-domain analysis of HRV [31]. The minimum, maximum, mean, and standard deviation values of the features and of the features selected by the wrapper method for the classifiers are given in Table 1.

3.4. Feature selection

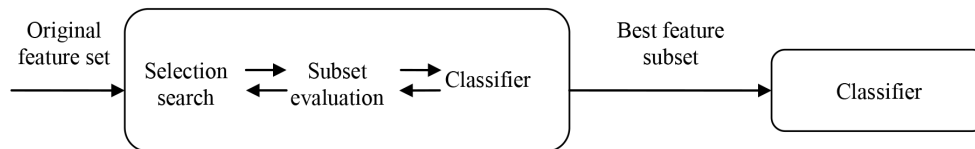
Noise or irrelevant features can cause undesirable difficulties due to the reduction of the accuracy of the classification methods. In order to eliminate this problem, feature selection methods were used for selecting a subset of relevant features to accomplish the best possible success with a specific learning algorithm. In machine learning and statistics, filter methods and wrapper methods are widely used feature selection approaches.

The filter methods search the original feature set given via a search algorithm and evaluate each possible subset by a simple filter method. They do not use any particular learning induction algorithm; hence the selected feature set can be used by different classifiers. They are not complex and are computationally faster than wrapper methods.

Similarly, wrapper methods use a search algorithm to search through the original feature set, but subset evaluation is achieved by using the classification algorithm itself. Then the best feature subset is selected based on the evaluation measure of classifier success (Figure 4). Since features were optimized for the classifier

Table 1. The HRV and EDR feature statistics and the selected features by wrapper.

Feature name	Min.	Max.	Average	Std. Dev.	Selected Features			
					k-NN	MLP	SVM	C4.5 DT
IQR	0.01	0.82	0.41	0.16	x	x	x	
MAD	0.01	0.94	0.20	0.07	x	x		x
Median	0.29	1.06	0.65	0.12	x	x		
SDNN	0.02	3.80	0.24	0.13		x	x	
RMSSD	0.02	3.79	0.24	0.13				
NN50v1	0	171	81.91	37.13	x	x	x	
NN50v2	0	93	46.32	19.19				
NN50v3	0	96	47.23	19.88	x	x		x
pNN50v1	0	99.37	47.22	20.35			x	x
pNN50v2	0	0.53	0.27	0.10	x	x		
pNN50v3	0	0.52	0.27	0.11		x		x
SDSD	0.01	5.19	0.17	0.17	x	x		
Mean	0.31	1.02	0.62	0.11	x	x	x	
HR	59.7	208.74	104.27	22.42	x	x		x
EDR IQR	0.07	4.09	0.95	0.58	x		x	
EDR MAD	0.07	2.82	0.54	0.31		x	x	x
EDR median	0.08	5.90	3.18	1.24			x	x

**Figure 4.** Wrapper approach for feature selection.

to be used, wrapper methods typically perform better. However, wrappers are more complex, require more computational time depending on the number of features to be optimized, and are susceptible to overfitting.

In this study, the wrapper-based feature selection method was used for the feature selection procedure. For evaluation of the subsets, each classifier with its default parameter value (except for k-NN) with the best first search method applying 10-fold cross-validation was employed using accuracy as the evaluation measure. For k-NN, a similar approach to that reported by Mendez et al. [13] was applied. First, k was set to 5 (arbitrarily) and the wrapper was employed to get the initial best subset. Thereafter, the first two distinctive features (IQR and median) given by the wrapper were selected, and k-NN was utilized with these features using different k values (1, 3, ..., 21) by employing Euclidean distance as a metric. The optimal k was chosen to be 15 based on the observed success. Finally, the wrapper was employed again by setting $k=15$ to obtain the final best subset. The best group of features evaluated for the classifiers were selected via the wrapper method and are given in Table 1.

3.5. k-Nearest neighbor (k-NN) classifier

k-NN is a nonparametric classifier that is considered a reference classifier by reason of its comparatively powerful method, and it has been employed in various science and engineering implementations [3]. k-NN is easy to apply since only the k , the number of nearest neighbors, should be determined as a parameter. k-NN is also a

lazy learning algorithm, as only the “k-nearest neighbors” are taken into consideration, and all calculation is delayed until classification. In k-NN classification, the input is the “k-nearest neighbor” and the output is a class membership to which the feature vector is assigned. The classification of a feature vector is handled by assigning it to the class of the most common among its “k-nearest neighbors”, which is achieved by a majority vote of its “k-nearest neighbors”. k-NN depends on a metric or a distance function such as Euclidean or that of Tanimoto [3]. Euclidean distance was used for its simplicity and accurate classification results. The general equation of Euclidean distance for N dimensional feature space is depicted as

$$d(a, b) = \sqrt{\sum_{i=1}^N (b_i - a_i)^2} = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \dots + (b_N - a_N)^2} \quad (6)$$

In Eq. (6), a ($a_1, a_2, a_3, \dots, a_N$) and b ($b_1, b_2, b_3, \dots, b_N$) are two N-dimensional vectors while $d(a, b)$ is the Euclidean distance between them [32]. The k is a user-defined integer. There exists an issue for binary classifications when the majority votes of the neighbors are tied. To avoid this issue, k should be defined as an odd number. The classification of a feature in one-dimensional feature space for binary classification is depicted in Figure 5 for three k-NN classifiers (1-NN, 3-NN, and 5-NN) based on the majority vote of the nearest neighbors selected according to the distance function. The decision of the 1-NN classifier ($k = 1$) for the feature O is simple since the nearest neighbor is a class Triangle member. For the 3-NN classifier ($k = 3$), the majority vote of 3 nearest neighbors (2 to 1) assigns the feature O to class Square. Finally, for the 5-NN classifier ($k = 5$), the feature O is assigned to class Triangle by the majority vote (3 to 2) of its 5 closest neighbors.

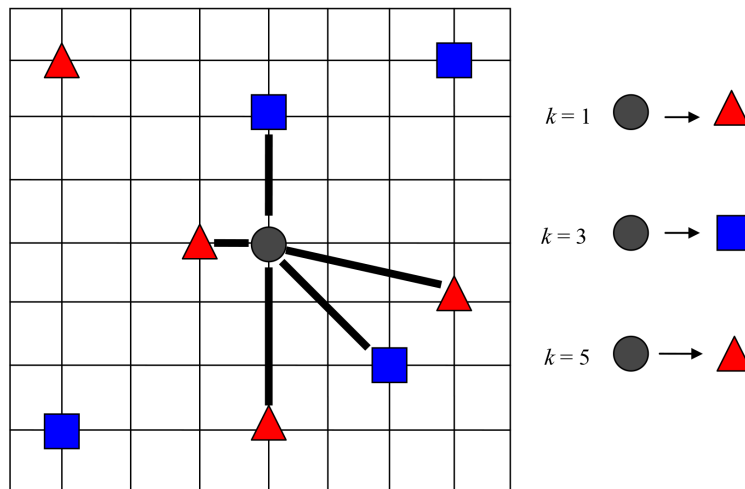


Figure 5. An example of k-NN classifier [32].

3.6. Optimal parameter calculation

In order to boost the classifier performance, the optimal parameters were calculated based on a trial-and-error approach applying 10-fold cross-validation. For this, the features selected by the wrapper method for each classifier were used to obtain the classifier performance for different parameters.

For k-NN, the parameter k 's success at different k values (1, 3, 5, ..., 21) was calculated by employing Euclidean distance as a metric. Figure 6 shows the curves of success at each k value. The optimal k was chosen to be 15 based on the observed success. For MLP, the optimal value of the parameter n (the number

of hidden nodes in the hidden layer) was calculated as 21 by adjusting different n values (1, 2, 3,..., 25) with iteration number of 500 epochs. For SVM, the gaussian radial basis function (RBF) kernel was employed, and the optimal values of the parameters C (the regularization parameter) and γ (gamma) were calculated as $C = 5$ and $\gamma = 0.1$ by adjusting different C values (0.1, 0.5, 1, 5, and 10) and γ values (0.1, 0.5, and 1). Finally, for C4.5 DT, the optimal values of the parameters CF (confidence factor) and MS (minimum number of split-off cases) were calculated as $CF = 0.1$ and $MS = 1$ by adjusting different CF values (0.01, 0.05, 0.1, 0.25, and 0.5) and MS values (1, 2, and 3).

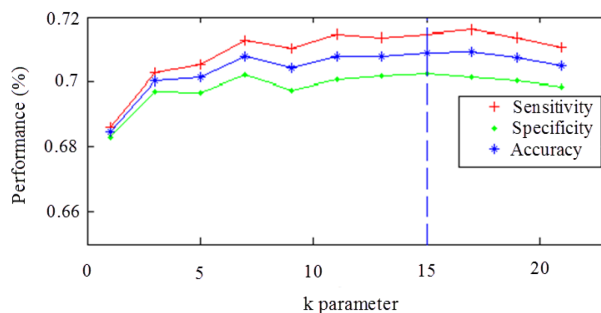


Figure 6. Classifier success at different k values for one parameter.

3.7. Classifier training

The training database including 12 recordings was taken from the apnea database of the DREAMS project. A total duration of 2984 min of the recording was scored as apneic (47.2%), while 3254 min was marked as normal (52.2%). For determining the success in training data, a 10-fold cross-validation technique was utilized. The performance for each fold for cross-validation is given in Table 2, and the training performance of the model for classification of apneic minutes is shown in Table 3 [33].

Table 2. The performance for each fold for cross-validation.

Fold #	Accuracy	Sensitivity	Specificity	AUC
1	0.76	0.75	0.78	0.84
2	0.78	0.80	0.79	0.86
3	0.80	0.83	0.79	0.87
4	0.77	0.79	0.77	0.85
5	0.77	0.78	0.77	0.84
6	0.76	0.78	0.76	0.84
7	0.76	0.76	0.78	0.85
8	0.77	0.78	0.78	0.84
9	0.74	0.76	0.75	0.83
10	0.76	0.79	0.75	0.83
Average	0.77	0.78	0.77	0.84

3.8. Testing

The test database including 35 recordings (15253 min) was provided by the Sleep Laboratory of Kocaeli University Hospital. As previously mentioned, 128 epochs containing high-amplitude artifacts caused by body movements were discarded, leaving 15,125 epochs to test. The models were tested with the test dataset, and

Table 3. Success of training data with selected features [32].

	Accuracy	Sensitivity	Specificity	AUC
k-NN	0.76	0.77	0.76	0.84
MLP	0.75	0.74	0.77	0.83
SVM	0.77	0.77	0.77	0.76
C4.5 DT	0.71	0.71	0.71	0.81

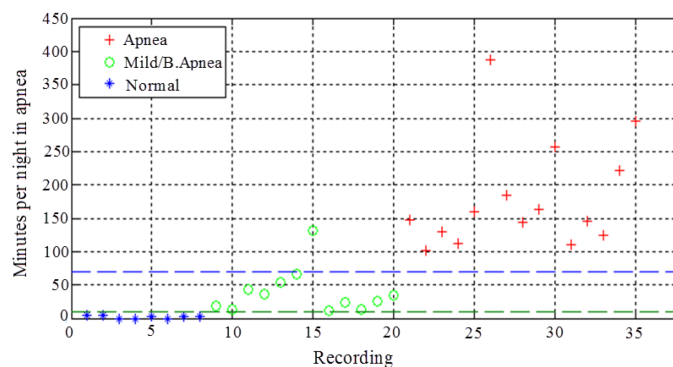
the number of apneic minutes for each subject was calculated. Two thresholds were used for class separation on minutes. The first threshold separating Class C (normal) also separates normal subjects from apnea classes (Class A and B), while the second threshold separates Class A (apnea) from Class B (mild/borderline apnea), distinguishing according to the severity estimate of OSAS.

k-NN was tested with 10 selected features and 3170 min (20.96%) were calculated as apneic for all datasets. The first threshold of 10 min for k-NN successfully separated all normal classes from apnea, while with the second threshold of 70 min, 26 apnea patients (Class A and B) out of 27 were correctly predicted. For MLP, with 12 selected features 3993 min (26.4%) were calculated as apneic; thresholds of 10 min and 100 min were applied and overall 29 patients out of 35 were correctly classified. For SVM, with 8 selected features 3175 min (20.99%) were calculated as apneic; thresholds of 15 min and 60 min were applied and overall 23 patients out of 35 were correctly classified. Finally, for C4.5 DT, with 8 selected features 2669 min (17.64%) were calculated as apneic; thresholds of 15 min and 100 min were applied, and overall 17 patients out of 35 were correctly classified.

4. Performance evaluation

The classification of the subjects in the test database was achieved by the diagnosis reports previously issued by physicians such as Class A (apnea): 15 subjects, Class B (mild/borderline apnea): 10 subjects, and Class C (normal): 8 subjects, according to the AHI values. The subject classification in the test dataset was performed by obtaining the minutes devoted to apnea and applying the aforementioned thresholds for each classifier.

A threshold of 10 min successfully separates all Class C (normal) subjects from the subjects with apnea (100% accuracy). Employing both thresholds for k-NN yields ‘one vs. all’ class prediction accuracy of 97.14% both for Class A and Class B, and 100% for Class C. The resulting overall prediction accuracy was 97.14%, corresponding to 34 subjects out of 35 on the test dataset. Class separation based on minutes per night achieved by the k-NN classifier is shown in Figure 7.

**Figure 7.** Class separation based on minutes per night.

The other classification models slightly underperformed; for example, normal subject separation from apnea with 97.1% accuracy was achieved both by MLP and SVM, and 86.1% by C4.5 DT, whereas the highest class prediction accuracy for Class A was obtained by MLP with 94.29%, and 91.43% for Class C. The success result of the classifiers on test dataset per subject class (one vs. all) is illustrated in Table 4.

Table 4. Success comparison of classifiers on test dataset per class.

Class	Classifier	Accuracy	Sensitivity	Specificity
Class A	k-NN	97.14	100	95.00
	MLP	94.29	100	90.00
	SVM	85.71	93.33	80.00
	C4.5 DT	82.86	66.67	95.00
	Average	90.00	90.00	90.00
Class B	k-NN	97.14	91.67	100
	MLP	91.43	75.00	100
	SVM	82.86	66.67	91.30
	C4.5 DT	80.00	75.00	82.61
	Average	87.86	77.08	93.48
Class C	k-NN	100	100	100
	MLP	100	97.14	88.89
	SVM	91.30	97.14	100
	C4.5 DT	82.61	86.11	77.78
	Average	93.48	95.10	91.67

5. Discussion

The significance of this work is that it proposes a pattern recognition system for automated apnea diagnosis based on HRV and EDR parameters obtained from single-lead ECG recordings and analyzed in the time domain using the k-NN classifier, which is simple and requires significantly less computational time compared with approaches using more complex classifiers and analyzing in the frequency domain.

Utilizing the ECG signal for apnea detection from multiple databases was encountered in the literature [5,14,16], but these studies are minimal in number. The majority of the related work consists of recordings from a single database [3,10–13,15,17,18] and mostly from the Physionet apnea-ECG database [34]. A common approach for sleep apnea detection from ECG is utilizing HRV. The related work based on features obtained from HRV [3,8–10,17,18] seems promising. Nevertheless, employing features from HRV and EDR together seems to be a good practice and provides better classification performance than using HRV features alone [11,14]. In the work by Penzel [35] in 2002, different algorithms to detect sleep apnea from ECG recordings, belonging to the Physionet apnea-ECG database [34] used in the annual Computers in Cardiology Challenge in 2000, were compared and it was reported that algorithms that used frequency-domain parameters (utilizing either FFT or DWT) performed the best. Moreover, the use of EDR parameters resulted in additional gains in performance. This is supported by the works in the literature that employed the parameters obtained from both HRV and EDR in the frequency domain [5,13-16]. Although using frequency-domain parameters seems more promising than time-domain parameters, it involves huge computational costs for pattern recognition systems, especially when it is combined with relatively more complex classifiers such as SVM and MLP. This is one of the main reasons why time-domain parameters were used in this study. On the other hand, the investigations in time-domain analysis dealing with features both from HRV and EDR are rare [14], as most of these works focus on using

features from HRV only [3,8–10]. In the present study, features extracted from HRV and EDR were analyzed in the time domain for the purpose of designing a simple, fast, and robust model for OSAS classification using ECG signals with only one lead that requires less computational time and is less complex. The k-NN classifier was focused on in this study because of its simplicity and robustness, and because it performs well with a low amount of data if it is trained with a distinctive dataset [13].

The HRV parameters employed in this work are the most commonly used cardiorespiratory features analyzed in the time domain, advised by the Task Force of ESC and the NASPE for time-domain analysis of HRV. The EDR signal was obtained by summing 50 ms of data on either side of each R peak, similar to [11,15]. The time scale of 1 min for epochs was chosen in this work similar to [5,10–15] for the purpose of comparing the results with the related works, and according to Chazal et al. [12], among the different epoch lengths (15 to 90 s), epoch lengths close to 60 s seem to be the most robust. The Physionet apnea-ECG database was most often used by related works [5,10–14,16–18] in which QRS complexes were provided; therefore, no R-wave detection procedure was required in those investigations. In this work, TEO was applied for R-wave detection similar to [28]. In addition, to evaluate the performance of TEO, the Pan–Tompkins algorithm used by [5,13] was also applied, and it was determined that the detected R-peak locations were alike.

For establishing an accurate pattern recognition system, feature extraction, feature selection, classifier optimization, training, and testing the classifier steps are followed sequentially. The time-domain HRV and EDR features were extracted, and to achieve the best possible success with the classifiers, the wrapper-based feature selection method was utilized for each employed classifier to select the best feature subsets using classifier accuracy as an evaluation measure. For boosting the classifiers' performance, the optimal parameters were calculated using a trial-and-error scheme by adjusting different values to the parameters according to accuracy, sensitivity, and specificity values. A 10-fold cross-validation technique was applied in the training procedure. Subsequently, classifiers were tested with the test dataset, and the number of apneic minutes for each subject was calculated. Feature selection, classifier optimization, and classification procedures consumed considerably less computation time for k-NN and C4.5 DT compared with MLP and SVM.

In the related works in which the k-NN classifier was utilized with time-domain features, Yılmaz et al. [3] studied HRV recordings of 17 subjects obtained from Gülhane Military Medical Academy (GATA) using IQR, median, and MAD in 30-s epochs, applying k-NN, SVM, and QDA. They achieved average accuracy rates of 79.5%, 87.2%, and 87.3%, respectively. Achieving such good accuracy rates with only three features is remarkable; using the same features in our training dataset, the accuracy rate for k-NN minute-based apnea detection was 72.9%, which is considerably high. However, in our study (with the 10 selected features obtained from both HRV and EDR), a higher accuracy rate of 76.67% was achieved. It is also stated that the mean and SDNN values were excluded due to the classification performance deteriorating since they are prone to be affected by the outliers. In order to analyze the impact of this issue, a model was constructed using the same selected features for k-NN excluding the mean (since SDNN is not among the selected features), and the training dataset was applied. As a result, an accuracy rate of 76.45% was achieved, whereas in our study, 0.22% higher accuracy was achieved when the mean was included. These differences may be because of the quality of the ECG signal used, additional features obtained from EDR, or the length of the epochs analyzed (30 s vs. 1 min). On the other hand, Isa et al. [10] analyzed the approaches of Yılmaz et al. [3] and 8 time-domain features (mean, SDNN, NN50v2, NN50v3, pNN50v2, pNN50v3, SDDSD, RMSSD) suggested by Chazal et al. [12], using k-NN, SVM, naïve Bayes, and QDA with the recordings from the Physionet apnea-ECG database [34] analyzed in 1-min epochs, and they reported that by employing 3 features proposed by Yılmaz et al. [3], 3.59%

higher classification accuracy was achieved. In their work, k-NN achieved accuracy of 75.63% in apneic minute detection and 77.78% in subject classification. In our work, in addition to the same features, NN50v1, pNN50v1, HR, EDR IQR, EDR median, and EDR MAD values were analyzed. The k-NN classifier achieved accuracy of 76.25% in apneic minute detection for the training dataset, and 97.14% in overall subject classification on the test dataset, which is much higher. These differences occurred as the result of employing different datasets and applying a feature selection procedure leading to the use of different features such as HR, NN50v1, and EDR IQR, which were among those selected for the k-NN classifier.

To summarize, k-NN, MLP, SVM, and C4.5 DT supervised learning classifiers with optimal ECG features were utilized in the time domain to obtain a better classification for automated sleep apnea diagnosis. Employing the thresholds to the model output with k-NN, overall 97.14% prediction accuracy for 34 subjects out of 35 on the test dataset was achieved. According to the obtained results, the k-NN classifier was determined to be a good choice for ECG-based apnea classification. Utilizing the wrapper-based features provided boosted the classification success with respect to accuracy, sensitivity, and specificity in both training and test processes. The proposed model supplied a classification rate of approximately 97.14%. Overall, 34 subjects out of 35 were correctly classified.

6. Conclusion and future works

PSG is recognized as the gold standard for diagnosis of sleep-related disorders, especially for OSAS. However, it is expensive and time-consuming. Consequently, more practical and less expensive solutions for apnea classification, like using only the ECG signal, which is applicable for Holter monitors, are appealing to researchers.

This study proposes a novel, simple, and fast pattern recognition system for automated apnea diagnosis based on HRV and EDR parameters using the k-NN classifier. By improving the classification success, this study can be used as a decision support system for the diagnosis of sleep apnea by specialists. The advantage of using only the ECG signal is to make the diagnosis process more practical and easy, requiring just a Holter device with no sleep laboratory necessary. Thus, this will result in eliminating the necessity of using sleep laboratories, which is expensive and involves discomfort for the patients when sleeping with a large number of sensors and cables. Patients can sleep at home in their own bed, feeling restful with only a Holter device. In addition, distinguishing OSAS patients from normal subjects will save time and provide the opportunity to physicians to focus on mild/severe OSAS patients. In future studies, different databases are planned to be employed with different machine learning techniques to achieve more reliable decision support systems in order to achieve quantification and classification of the severity of sleep apnea.

Acknowledgement

We thank Kocaeli University Sleep Laboratory for sharing their sleep data with us.

References

- [1] Wolk R, Gami AS, Garcia-Touchard A, Somers VK. Sleep and cardiovascular disease. *Curr Prob Cardiology* 2005; 30: 625-662.
- [2] Morgenthaler TI, Kagramanov V, Hanak V, Decker PA. Complex sleep apnea syndrome: is it a unique clinical syndrome? *Sleep* 2006; 29: 1203-1209.
- [3] Yılmaz B, Asyalı MH, Arıkan E, Yetkin S, Özgen F. Sleep stage and obstructive apneic epoch classification using single-lead ECG. *Biomed Eng Online* 2010; 9: 1-14.

- [4] Otero A, Félix P, Barrob S, Zamarrón C. A structural knowledge-based proposal for the identification and characterization of apnoea episodes. *Appl Soft Comput* 2011; 12: 516-526.
- [5] Mendez MO, Bianchi AM, Matteucci M, Cerutti S, Penzel T. Sleep apnea screening by autoregressive models from a single ECG lead. *IEEE T Bio-Med Eng* 2009; 56: 2838-2849.
- [6] Guilleminault C, Connolly S, Winkler R, Melvin K, Tilkian A. Cyclical variation of the heart rate in sleep apnea syndrome. *Lancet* 1984; 21: 126-131.
- [7] Timuş O, Kiyak E. Optimizing MLP classifier and ECG features for sleep apnea detection. *Journal of Naval Science and Engineering* 2015; 11: 1-18.
- [8] Roche F, Gaspoz J M, Court-Fortune I, Minini P, Pichot V, Duverney D, Costes F, Lacour JR, Barthelemy JC. Screening of obstructive sleep apnea syndrome by heart rate variability analysis. *Circulation* 1999; 100: 1411-1415.
- [9] Roche F, Pichot V, Sforza E, Court-Fortune I, Duverney D, Costes F, Garet M, Barthelemy JC. Predicting sleep apnea syndrome from heart period: a time-frequency wavelet analysis. *Eur Respir J* 2003; 22: 937-942.
- [10] Isa S, Fanany MI, Jatmiko W, Murini A. Feature and model selection on automatic sleep apnea detection using ECG. In: 2010 International Conference on Computer Science and Information Systems; 20–23 Nov 2010; Bali, Indonesia. pp. 357-362.
- [11] Chazal P, Heneghan C, Sheridan E, Reilly RB, Nolan P, O'Malley M. Automated processing of the single lead electrocardiogram for the detection of obstructive sleep apnea. *IEEE T Bio-Med Eng* 2003; 50: 686-696.
- [12] Chazal P, Penzel T, Heneghan C. Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram. *Physiol Meas* 2004; 25: 967-983.
- [13] Mendez MO, Ruini DD, Villantrieri OP, Matteucci M, Penzel T, Cerutti S, Bianchi AM. Detection of sleep apnea from surface ECG based on features extracted by an autoregressive model. In: IEEE 2007 29th International Engineering in Medicine and Biology Society Conference; 23–26 Aug 2007; Lyon, France. New York, NY, USA: IEEE. pp. 6105-6108.
- [14] Khandoker AH, Palaniswami M, Karmakar M. Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings. *IEEE T Inf Technol B* 2009; 13: 37-48.
- [15] Yıldız A, Akın M, Poyraz M. An expert system for automated recognition of patients with obstructive sleep apnea using electrocardiogram recordings. *Expert Syst Appl* 2011; 38: 12880-12890.
- [16] Chen L, Zhang X, Song C. An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram. *IEEE T Auto Sci Eng* 2015; 12: 106-115.
- [17] Gürüler H, Şahin M, Ferikoğlu A. Feature selection on single-lead ECG for obstructive sleep apnea diagnosis, *Turk J Elec Eng & Comp Sci* 2014; 22: 465-478.
- [18] Babaeizadeh S, White DP, Pittman SD, Zhou SH. Automatic detection and quantification of sleep apnea using heart rate variability. *J Electrocardiol* 2010; 43: 535-541.
- [19] Iber C, Ancoli-Israel S, Chesson A, Quan SF. The AASM Manual for the Scoring of Sleep and Associated Events. Westchester, IL, USA: American Academy of Sleep Medicine, 2007.
- [20] Devuyst S, Dutoit D, Kerkhofs M, The DREAMS database, University of MONS - TCTS Laboratory and Université Libre de Bruxelles - CHU de Charleroi Sleep Laboratory.
- [21] Shoelson B. A simple file reader for European Data Formatted (EDF-) files. Mathworks Matlab Center, 2013.
- [22] Kranjec J, Beguš S, Geršak G, Drnovšek J. Non-contact heart rate and heart rate variability measurements: a review. *Biomed Signal Process Control* 2014; 13: 102-112.
- [23] Yakut O, Solak S, Bolat ED. Measuring ECG signal using e-health sensor platform. In: 2014 International Conference on Chemistry, Biomedical and Environment Engineering; 7–8 Oct 2014; Antalya, Turkey. pp. 71-75.
- [24] Li C, Zheng C, Tai C. Detection of ECG characteristic points using wavelet transforms. *IEEE T Bio-Med Eng* 1995; 42: 21-28.

- [25] Martinez JP, Almeida R, Olmos S, Rocha AP, Laguna P. A wavelet based ECG delineator: evaluation on standard databases. *IEEE T Bio-Med Eng* 2004; 51: 570-581.
- [26] Sahambi JS, Tandon SN, Bhatt RKP. Using wavelet transforms for ECG characterization. *IEEE Eng Med Biol* 1997; 16: 77-83.
- [27] Saxena SC, Kumar V, Hamde ST. Feature extraction from ECG signals using wavelet transforms for disease diagnostics. *Int J Syst Sci* 2002; 33: 1073-1085.
- [28] Koçak O, Bayrak T, Erdamar A, Özparlak L, Telatar Z, Eroğul O. Automated detection and classification of sleep apnea types using electrocardiogram (ECG) and electroencephalogram (EEG) features. *Adv Electrocard-Clin Appl* 2012: 211-230.
- [29] Hamila R, Astola J, Alaya Cheikh F, Gabbouj M, Renfors M. Teager energy and the ambiguity function. *IEEE T Signal Proces* 1999; 47: 260-262.
- [30] Kaiser JF. Some useful properties of teager's energy operators. In: *IEEE 1993 Acoustics, Speech And Signal Processing Conference*; 27-30 April 1993; Minneapolis, MN, USA: IEEE. pp: 149-152.
- [31] Malik M. Heart rate variability, standards of measurement, physiological interpretation, and clinical use, task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Eur Heart J* 1996; 17: 354-381.
- [32] Laboratory 8. Classification and Vowel Recognition, The University of Michigan.
- [33] Timuş O. Sleep respiration disorders diagnosis and classification utilizing soft computing algorithms. PhD, Kocaeli University, Kocaeli, Turkey, 2015.
- [34] Penzel T, Moody GB, Mark RG, Goldberg AL, Peter JH. The apnea-ECG database. *IEEE Comput Cardio* 2000; 27: 255-258.
- [35] Penzel T, Mcnames J, de Chazal P, Raymond B, Murray A, Moody G. Systematic comparison of different algorithms for apnea detection based on electrocardiogram recordings. *Med Biol Eng Comput* 2002; 40: 402-407.