

Development of hybrid artificial intelligence based automatic sleep/awake detection

ISSN 1751-8822
Received on 21st January 2019
Revised 28th July 2019
Accepted on 9th October 2019
E-First on 13th February 2020
doi: 10.1049/iet-smt.2019.0034
www.ietdl.org

Mehmet Recep Bozkurt¹, Muhammed Kürşad Uçar¹ ✉, Ferda Bozkurt², Cahit Bilgin³

¹Faculty of Engineering, Electrical-Electronics Engineering, Sakarya University, Sakarya, Turkey

²Vocational School of Adapazarı, Computer Programming, Sakarya University of Applied Sciences, Sakarya, Turkey

³Faculty of Medicine, Sakarya University, Sakarya, Turkey

✉ E-mail: mucar@sakarya.edu.tr

Abstract: Background and Objective: Obstructive Sleep Apnea is a disease that causes respiratory arrest in sleep and reduces sleep quality. The diagnosis of the disease is made by the physician in two stages by examining the patient records taken with the polysomnography device. Because of the negative aspects of this process, new diagnostic processes and devices are needed. In this article, a new approach to sleep staging, which is one of the diagnostic steps of the disease, was proposed. An artificial intelligence-based sleep/awake system detection was developed for sleep staging processing. Photoplethysmography (PPG) signal and heart rate variable (HRV) were used in the study. PPG records taken from patient and control groups were cleaned by the digital filter. The HRV parameter was then derived from the PPG signal. Then, 40 features from HRV signal and 46 features from PPG signal were extracted. The extracted features were classified by reduced machine learning techniques with *F*-score feature selection method. In order to evaluate the performances of the classifiers, the sensitivity and specificity values, the accuracy rates for each class were computed in the test set and receiver operating characteristic curve prepared. In addition, area under the curve (AUC), Kappa coefficient and *F*-score were calculated. According to the results obtained, the system can be realised with 91.09% accuracy rate using 11 PPG and HRV and with 90.01% accuracy rate using 14 HRV features. These success rates are quite enough for the system to work. When all these values are taken into consideration, it is possible to realise a practical sleep/awake detection system. This article suggests that the PPG signal can be used to diagnose obstructive sleep apnea by processing with artificial intelligence and signal processing techniques.

1 Introduction

Reduced breathing quality in sleep causes obstructive sleep apnea (OSA). Patient records are collected with the polysomnography device for diagnosis. The diagnosis is made by the physician according to the diagnostic rules of the American Academy of Sleep Medicine [1]. Sleep staging is used to measure the amount of time the patient spends sleeping, and respiratory scoring is used to detect the number and duration of abnormal respiratory events that occur during sleep.

Sleep staging can be carried out by electroencephalography (EEG), electrocardiogram (EOG) and electromyogram (EMG) signals with a minimum of 16 channels and only by a specialist physician [1]. Respiratory scoring is done by oral-nasal airflow signal, blood oxygen saturation, thorax, and abdominal respiratory motion signals [1]. As a result of these steps, the number of abnormal respiratory events is divided by the duration of sleep and the apnea-hypopnea index (AHI) is calculated. If $AHI < 5$, the individual is considered normal. If $AHI > 5$, the individual has OSA and the level of the disease is determined by looking at the AHI value. The diagnostic process is vital in diagnosing sleep staging and respiratory scoring when examined in general terms. There are many disadvantages of the system, such as the troublesome diagnosis process, inadequate use of the device at home, the patient being away from the sleeping environment due to the excess of electrodes, necessity of the technician to connect the system to the patient and necessity of examining the data only by a specialist doctor. These disadvantages of the system triggered the development of new systems [2, 3]. The objective of this study is to develop an artificial intelligence-based hybrid system with a high accuracy rate in OSA diagnosis and an alternative to the PSG process with fewer sensors.

With sleep staging, the patient's sleep is analysed. The analysis is performed according to the sleep staging rules of the AASM [1]. For analysis, signals from the eyes are used [1]. The sleeping

period is divided into pieces. Each piece is 30 s. Each piece is named according to the characteristics of the signals. These are W: awake, N1: stage 1, N2: stage 2, N3: stage 3 and rapid eye movement (REM) [1]. The purpose of this analysis is to identify the parts of the sleep staging where the patient is asleep. In OSA diagnosis, sleep staging is a prerequisite for the respiratory scoring process to be performed.

In the literature attempted to perform sleep staging with the classification of EEG signals [4–9]. There are different studies carried out to make sleep staging more practical [5, 8, 10–12]. Heart rate variable (HRV) is the most used of these methods [8, 10, 12, 13]. Since HRV can be derived from photoplethysmography (PPG) signal as well as ECG, it is thought that there may be a relationship between PPG and HRV and sleep stages [14]. In addition, medical signal and machine-learning-based sleep phases are also detected [15–17]. Neural networks have been used in some of these studies [18]. Even commercial automatic sleep staging systems (e.g. ZMachine®) have been developed [19].

PPG is a signal that can be measured from any part of the body and contains some information about the body [14]. HRV is a parameter that can be derived from the ECG or PPG signal, relaxation and sleep by the autonomic nervous system during a 24-hour period [13, 20].

The aim of the study is to develop a non-invasive, portable, fast and reliable system for sleep staging in the diagnosis of OSA.

The difference of this study from the literature is the use of PPG signal, artificial intelligence algorithms as well as signal processing techniques in conjunction to analyse sleep staging. [4–8].

In this study, sleep/awake detection process was processed by PPG signal processing techniques and then the sleep/awake states were tried to be determined by artificial intelligence algorithms depending on the PPG signal.

PPG can be obtained more quickly than EEG, EOG, EMG, and ECG signals used in sleep staging. The PPG measurement device gives the patient less disturbance than other signals signal

acquisition. It is also a reliable signal such as EEG, EOG, EMG, and ECG. Therefore, PPG was preferred in this study.

The flow of the article is as follows; In the Section 2, the database used in the study is defined, signal processing and feature extraction steps are explained for the PPG signal and the machine learning techniques used are introduced. Simulation results are interpreted in the Section 3, while the results of the study were interpreted in the Section 4.

2 Methods

In this study, the signal processing steps are performed according to the stages shown in Fig. 1. According to these steps, a database of PPG records belonging to the individuals was created. Subsequently, digital filtering was performed to remove noise on the PPG signals and HRV was derived from the PPG signals. Subsequently, feature extraction from PPG signal and HRV was performed. According to the relation of the extracted features with apnea, the features that are compatible with the feature selection algorithm are selected. Then, they were classified by artificial intelligence algorithms and system performance was tested.

2.1 Signal acquisition

Within the scope of this study, the records were recorded with SOMNOscreen Plus PSG device at the Chest Diseases, Sakarya Hendek State Hospital. In total, 33 channel signals of ten patients were recorded. However, the study was carried out with PPG signals from the finger with a sampling frequency of 128 Hz. Thirty three channel signals consist of the following: 25 channels are for EEG/EOG, six channels are for EMG, one channel is for ECG, one reference electrode. In addition, body position, breathing effort, ambient light, movement, patient marker information are being collected with the patient's channel. All recordings are saved simultaneously. A 10/20 electrode arrangement was used for EEG measurements. Individual demographic information is shown in Table 1.

Patients are admitted to Sakarya Hendek State Hospital every day. The selected ten patients were randomly selected from patients who arrived here without a certain period. The patients were

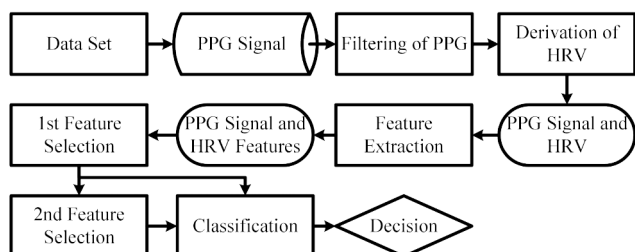


Fig. 1 General signal processing flow diagram

informed about the studies and consent was obtained. Data permission and ethics committee report were obtained for the study. For detailed information, see Section 7.

The records were reviewed by a specialist physician. The sleep tags used are W, N1, N2, N3, and REM. N1, N2, N3, and REM were labelled with a single label as Sleep (S) since it was sufficient to determine sleep and awake states for respiratory scoring. Thus, two sleeping labels are used in total, including W and S. Table 1 shows the number of W and S stages taken from each individual. Each stage represents 30 s of epochs. A total of 1482 awake epochs and 6953 sleep epochs were obtained from all subjects. Epok is a recording piece that contains a certain period. The label of this part is given by the doctor (sleep or awake).

2.2 Signal pre-processing

The noise on the PPG signals was cleaned with two different filters sequentially. The first filter is a 0.1–20 Hz IIR-Chebyshev type II bandpass filter and the second filter is the moving average filter. In order to clean the noise on the PPG signals, 0.1–20 Hz IIR-Chebyshev type II band-pass and moving average filter were applied. The PPG signal carries information between 0 and 20 Hz [21]. The reason that the designed filter started at 0.1 Hz was to be able to eliminate the direct current component on the signal [22]. Then, HRV parameters are derived from the cleaned PPG signal. When each HRV parameter was calculated from the PPG signal, the peak points of the PPG signal were determined. The $N - 1$ HRV parameter was calculated for the determined N peak. An HRV parameter corresponds to the time between two successive peaks of the PPG.

For example, from the 5 s PPG signal, let us suppose that $N = 4$ local maximum points, A_{P1}, A_{P2}, A_{P3} and A_{P4} are identified. In this case, $N - 1 = 3$ HRV parameters A_{H1}, A_{H2} and A_{H3} can be calculated. For example, x, y axis coordinates of A_{P1} are 105.79.8726 while $A_{P2} = 105$. The point where A_{P2} is located is 246.80.5023 while $A_{P2} = 246$. In this case, a sample calculation for A_{H1} is shown in (1). These operations were applied to every PPG epoch of 30 s.

$$A_{H(i)} = \frac{A_{P(i+1)} - A_{P(i)}}{fs} \rightarrow$$

$$\rightarrow A_{H1} = \frac{A_{P2} - A_{P1}}{fs} = \frac{246 - 105}{128} = 1.1016 \text{ s} \quad (1)$$

2.3 Feature extraction

A total of 46 features were extracted from the PPG signal. Out of 46 features, 10 of them are frequency domain, 36 of them are time domain. A total of 40 features have been extracted from the HRV signal. Out of 40 features, 10 of them are frequency domain and 30 of them are time domain. Some features of PPG and HRV are

Table 1 Demographic information and data distributions of apnea and control groups

Information	Female $n_1 = 5$	Male $n_2 = 5$	All individuals $n = n_1 \pm n_2 = 10$
age, y	59 ± 5	53 ± 11.31	56 ± 8.79
weight, kg	103.6 ± 10	102.02 ± 6.8	102.81 ± 8.28
height, cm	162 ± 3	173 ± 2.83	167.5 ± 6.43
body mass index (BMI), kg/m ²	39.46 ± 3	34.14 ± 3.06	36.8 ± 4.05
AHI	9.52 ± 6	24.38 ± 13.21	16.95 ± 12.52

Distribution of sleep epochs

Individual number	1	2	3	4	5	6	7	12	19	27	Total
Sex	Female	Male	Male	Male	Male	Male	Female	Female	Female	Female	
awake (epoch)	43	103	139	90	566	11	230	76	127	97	1482
sleep (epoch)	835	749	680	751	256	838	606	767	659	812	6953
total recording time, h	7.32	7.10	6.83	7.01	6.85	7.08	6.97	7.03	6.55	7.58	70.29

Each sleeping epoch contains 30 s of recording.

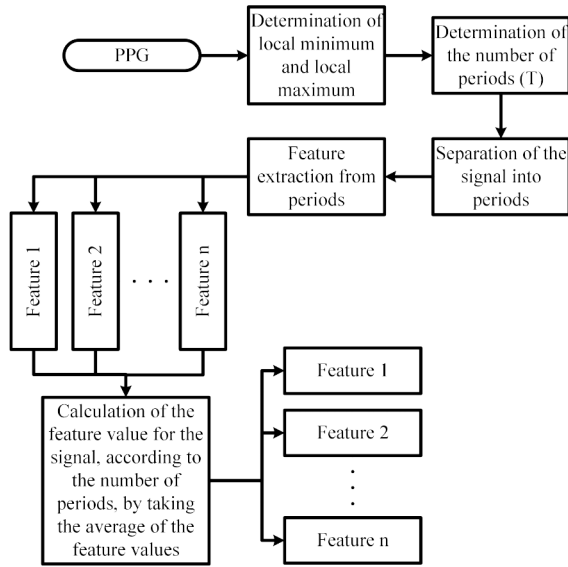


Fig. 2 Feature extraction flow diagram

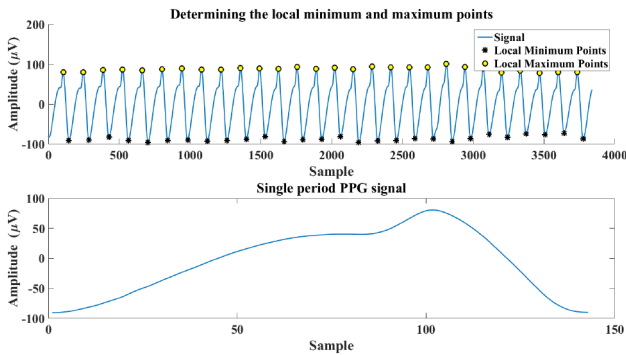


Fig. 3 Determining the local minimum and maximum points, and a single period signal

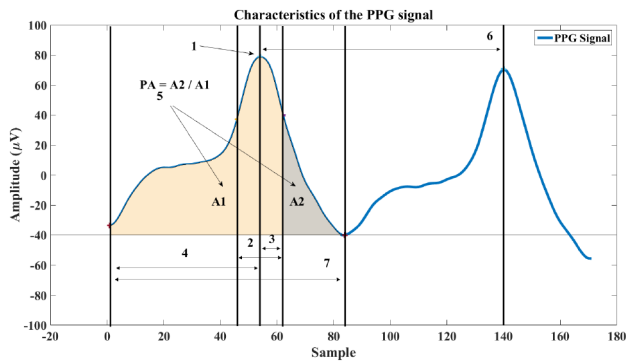


Fig. 4 Features of the PPG signal

common. Therefore, in this article, the narration is common in some places.

The feature extraction process is shown in Fig. 2. Accordingly, the minimum and maximum points of the signal are determined first. According to these points, the signal is divided into periods. The start and end points of the signal are the minimum points. T is the number of periods, $LOCMIN$ is the local minimum number and $T = LOCMIN - 1$.

Fig. 3 shows an example PPG signal. The first signal is 30 s. The second signal is the periodically determined signal. The number of periods of the signal is $T = LOCMIN - 1 = 28 - 1 = 27$.

The desired characteristics were extracted from each period of the PPG signal whose periods were detected and recorded as the relevant epoch feature by the average. In this case, it is ensured that the properties were obtained with minimum error rate. For example, when the shape factor feature was calculated, the shape

factor values for each period obtained were calculated separately, and the only shape factor value of the epoch was obtained by the average of these values is taken. These steps were done again for every feature.

In the study, seven characteristic features were extracted from the PPG signal (Fig. 4). The systolic peak value is the first feature. The second feature is the value in seconds between the points where the peak of the PPG signal is equal to the half of the systolic peak amplitude from the right and left of the peak. The third property represents the time in the place marked as 3 on the figure. The time between the start of the signal and the systolic peak is the fourth feature. The ratio of the fields shown as A1 and A2 is the sixth feature ($PA = A2/A1$). The time between two systolic peaks is the sixth feature. A period of the signal is the seventh feature.

Table 2 shows the PPG and HRV properties with their formulas and numbers. Features marked with ‘*’ are calculated using the MATLAB library [23]. The x in the formulas represents the signal. If there is a ‘—’ in the feature number section, that feature is not calculated.

In addition to the statistical characteristics, the energy quantities in the sub frequency bands of HRV and PPG signal were determined and used as frequency domain characteristics. When the frequency domain characteristics are extracted, the low frequency (LF) bands belonging to the signals are extracted first.

The PPG signal has three sub-frequency bands, these are (i) 0.15–0.6 Hz high-frequency (HF) band, (ii) 0.09–0.15 Hz medium frequency band, (iii) 0.04–0.15 Hz LF band [11, 24]. HRV has three sub-frequency bands, these are (i) 0.15–0.4 Hz HF band, (ii) 0.04–0.15 Hz LF band, (iii) 0.0033–0.04 Hz very low frequency band [11, 24].

The energies are calculated after separating PPG and HRV into lower frequency bands (Denklem 2).

$$E = \sum_{i=-\infty}^{+\infty} |x[i]|^2 \quad (2)$$

The calculated energies are indicated by the symbols E_{PPG} , E_{PPGLF} , E_{PPGMF} , E_{PPGHF} , E_{HRV} , E_{HRV_VLF} , E_{HRV_LF} , E_{HRV_HF} (Table 3).

2.4 F-score

F -score is a mathematical algorithm that is used to reveal distinguishing features between groups. The process steps are as shown in Fig. 5. An F -score value is calculated for each feature (Denklem 3) [25]. The F -score threshold value is calculated by taking the average of all F -score values. If the F -score value of a feature is greater than the threshold value, that feature is selected. The F -score value of the feature is directly proportional to the degree of discrimination.

The parameter definitions in (3) are as follows: (i) $\mathbf{x}_{k,i}$ is the feature vector, (ii) n_a is the number of elements in class a, (iii) n_b is the number of elements in class b, (iv) $m = n_a + n_b$, which is total number of elements, (v) $k = 1, 2, \dots, m$, (vi) i is the feature number, (vii) the average value of the $\bar{x}_i^{(a)}$ attribute, (viii) $\bar{x}_i^{(a)}$ is the average value of the i attribute in class a, (ix) $\bar{x}_i^{(b)}$ is the average value of the i attribute in class b, (x) $\mathbf{x}_{k,i}^{(a)}$ in class a, is the i feature of the k element, (xi) $\mathbf{x}_{k,i}^{(b)}$ in class b, is the i feature of the k element.

$$F(i)$$

$$= \frac{(\bar{x}_i^{(a)} - \bar{x}_i)^2 + (\bar{x}_i^{(b)} - \bar{x}_i)^2}{(1/(n_a - 1)) \sum_{k=1}^{n_a} (\mathbf{x}_{k,i}^{(a)} - \bar{x}_i^{(a)})^2 + (1/(n_b - 1)) \sum_{k=1}^{n_b} (\mathbf{x}_{k,i}^{(b)} - \bar{x}_i^{(b)})^2} \quad (3)$$

F -score feature selection algorithm determines the power of each feature independently of the other features. Evaluation, implementation, and writing code is very easy. It works faster than other algorithms. Because of these advantages, F -score feature selection algorithm was used in the study [25].

Table 2 Formulas for PPG and HRV features in time domain

PPG feature number	Feature	Formula	HRV feature number
8	mean	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n}(x_1 + \dots + x_n)$	1
9	standard deviation	$S = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$	2
10	average curve length	$CL = \frac{1}{n} \sum_{i=2}^n x_i - x_{i-1} $	3
11	average energy	$E = \frac{1}{n} \sum_{i=1}^n x_i^2$	4
12	average Teager energy	$TE = \frac{1}{n} \sum_{i=3}^n (x_{i-1}^2 - x_i x_{i-2})$	5
13	Hjort parameters – activity	$A = S^2$	6
14	Hjort parameters – mobility	$M = S_1^2 / S^2$	7
15	Hjort parameters – complexity	$C = \sqrt{(S_2^2 / S_1^2)^2 - (S_1^2 / S^2)^2}$	8
—	maximum ^a	$x_{\max} = \max(x_i)$	9
16	skewness	$x_{\text{ske}} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)S^3}$	10
17	Kurtosis	$x_{\text{kur}} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(n-1)S^4}$	11
18	shape factor	$SF = X_{\text{rms}} / \left(\frac{1}{n} \sum_{i=1}^n \sqrt{ x_i } \right)$	12
19	minimum ^a	$x_{\min} = \min(x_i)$	13
20	root mean squared value	$X_{\text{rms}} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i ^2}$	14
21	singular value decomposition ^a	SVD = svd(x)	15
22	median	$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & x \text{ odd} \\ \frac{1}{2}(x_{n/2} + x_{(n/2)+1}) & x \text{ even} \end{cases}$	16
23	geometric mean	$G = \sqrt[n]{x_1 + \dots + x_n}$	17
24	harmonic mean	$H = n / \left(\frac{1}{x_1} + \dots + \frac{1}{x_n} \right)$	18
25	25% trimmed mean ^a	$T25 = \text{trimmean}(x, 25)$	19
26	50% trimmed mean ^a	$T50 = \text{trimmean}(x, 50)$	20
27	range	$R = x_n - x_1$	21
28	interquartile range ^a	$IQR = \text{iqr}(x)$	22
29	mean or median AD ^a	$MAD = \text{mad}(x)$	23
30	moment, central moments ^a	$CM = \text{moment}(x, 10)$	24
31	coefficient of variation	$DK = (S/\bar{x})100$	25
	normality test p^a	$[p, h] = \text{kstest}(x)$	26
	normality test $h_{1,0}^a$		27
32	sign test p^a	$[p, h] = \text{signtest}(x)$	28
33	sign test $h_{1,0}^a$		29
34	standard error	$S_{\bar{x}} = S/\sqrt{n}$	30
35	YMaks	YMaks	—
36	YMin	YMin	—

YMaks, number of local maximum in epochs; YMin, number of local minimum in epochs; AD, absolute deviation.

^aThe PPG and HRV features extracted with MATLAB commands.

2.5 Classification stage

The purpose of the classification process was to perform sleep/awake detection based on machine learning using the PPG signal and HRV features. The classification process was carried out in two steps shown in Fig. 6. In the first stage, the properties were classified without any treatment. Then, the feature was applied to the F -score feature selection algorithm twice and classified in each step.

Four different machine learning techniques were used for the classification process. These techniques were as follows: k nearest

neighbours (kNNs) algorithm, probabilistic artificial neural networks (PNN), multilayer feedforward artificial neural networks (MLFFNN) and support vector machines (SVMs). In addition, the ensemble classifier working with the common decision of all these classifiers was used.

2.5.1 kNN classifier: kNN is one of the trainer classifier algorithms [26]. The important thing in the method is that the properties of each class have been determined beforehand. The performance of the method is influenced by the closest neighbour

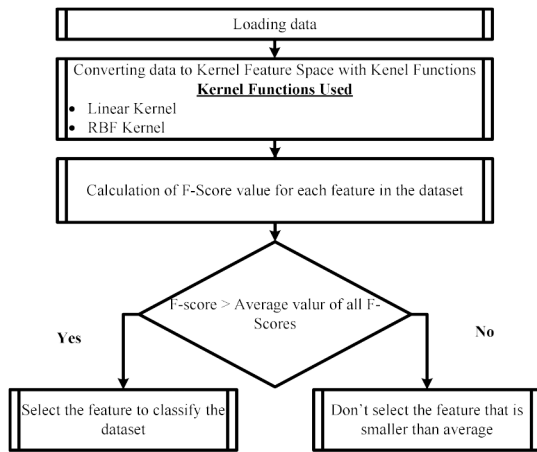


Fig. 5 Kernel F-score feature selection steps

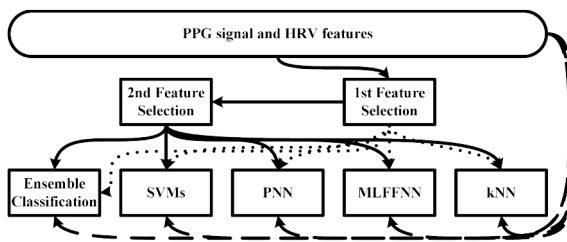


Fig. 6 Classification flow diagram

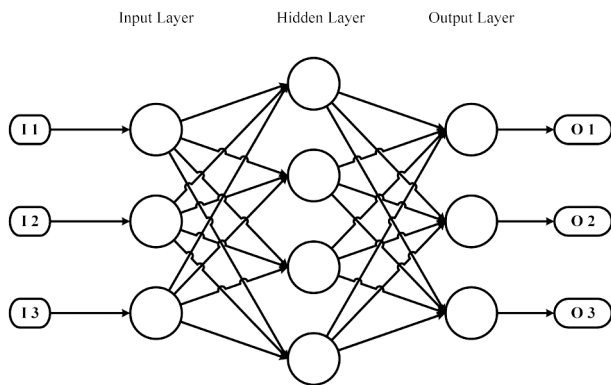


Fig. 7 General network structure for MLFFNN and PNN

number, the threshold value, the measure of similarity, and the criteria in which the data in the learning set are sufficient. k values are initially selected. The large selection of k value may lead to the aggregation of unequal data groups. In practice, k is generally preferred to be 3, 5 or 7 [27].

kNN classifier works in the following way; at the beginning, k is determined. The values of $k = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ were used in this study. Then the distance between the class tag unknown data and the known data is calculated. Distance calculations can be made with various distance calculation formulas. In total, 11 different distance calculation formulas were used in this study, these formulas were: Chebychev, Cityblock, Correlation, Cosine, Euclidean, Hamming, Jaccard, Mahalanobis, Minkowski, Seucclidean, Spearman. The kNNs are determined according to the distance formulas. The majority is determined by k neighbours. The majority label is set as the new label of the unknown label. Ten different k values and 11 different distance calculation formulas were used in the study. In this way, the kNN network of 110 different structures was established to classify a data set and the most efficient result was determined.

2.5.2 Multilayer feedforward artificial neural networks: Artificial neural networks are a classification tool created by combining artificial neural cells [28]. In this structure, the data moves in one direction. [29]. The network consists of three

Table 3 Features of frequency domain

PPG feature number	Formula	Feature	Formula	HRV feature number
37	E_{PPG}	energy	E_{HRV}	31
—	—	—	$E_{HRV_{VLF}}$	32
38	$E_{PPG_{LF}}$	—	$E_{HRV_{LF}}$	33
39	$E_{PPG_{MF}}$	—	—	—
40	$E_{PPG_{HF}}$	—	$E_{HRV_{HF}}$	34
41	$E_{PPG_{LF}}/E_{PPG}$	—	$E_{HRV_{VLF}}/E_{HRV}$	35
42	$E_{PPG_{MF}}/E_{PPG}$	—	$E_{HRV_{LF}}/E_{HRV}$	36
43	$E_{PPG_{HF}}/E_{PPG}$	—	$E_{HRV_{HF}}/E_{HRV}$	37
44	$E_{PPG_{LF}}/E_{PPG_{MF}}$	—	$E_{HRV_{VLF}}/E_{HRV_{LF}}$	38
45	$E_{PPG_{LF}}/E_{PPG_{HF}}$	—	$E_{HRV_{VLF}}/E_{HRV_{HF}}$	39
46	$E_{PPG_{MF}}/E_{PPG_{HF}}$	—	$E_{HRV_{LF}}/E_{HRV_{HF}}$	40

Table 4 Network operation parameters

Training algorithm	NN	R
Levenberg–Marquardt	trainlm	1 10
BFGS quasi-Newton	trainbfg	2
resilient backpropagation	trainrp	3
scaled conjugate gradient	trainscg	4
conjugate gradient with Powell/Beale restarts	traincgb	5
Fletcher–Powell conjugate gradient	traincgf	6
Polak–Ribière conjugate gradient	traincgp	7
one step secant	trainoss	...
variable learning rate gradient descent	trainidx	100

NN, number of neurons; R, repetition.

layers (Fig. 7), these are the input layer, the hidden layer, and the output layer.

There are several initial parameters for the MLFFNN classifier (Table 4). Nine different training algorithms and neuron number (1–100) parameters were used for MLFFNN in the study, for a second time. The MLFFNN network has been run for ten times with the same parameters to obtain better results in operation. For example, in the classification of PPG data, a total of $9 \times 100 \times 10 = 9000$ different networks were run and the best network results were chosen. Considering that there are nine different data sets, a total of 81,000 different MLFFNN networks were created in the study.

2.5.3 Probabilistic neural networks: PNN is a Bayesian-based general classifier [30]. PNN is a network that is based on the generalisation and takes into account all the points in the classification process. For classification, the distance is calculated from the evaluated point to each other point. The distance is calculated by the kernel function.

The PNN network structure is similar to forward feed network structures. In Fig. 7 the general network structure is shown. The number of neurons in the input layer is equal to the number of features. The number of features that can be used as a network entry in the study is used as the number of selected features (as 4, 5, 11, 16 and 28 numbers) as summarised in Table 5. There are two hidden layers in the network structure. The number of neurons in the second layer is 2. Since the study has two different output values (apnea/control), the output layer has two neurons.

For the PNN classifier, only the spread initialisation parameter can be intervened. As the spread parameter approaches zero, the network begins to behave like the nearest neighbour classifier [31]. As this value moves away from zero, the classifier classifies several vectors, which separate the data from each other [31]. This value is designed between 0.001 and 5 with a range of 0.001 steps in a total of 5000 different spread parameters. At the end of the study, the best performing network parameters and performance criteria were calculated.

Table 5 Feature selection results with *F*-score for sleep/awake detection

Signal	(1) Feature selection										(2) Feature selection					
	FC SFC					SFN					SFC	SFN				
PPG	46	21	2	3	9	10	11	12	13	6	10	16	6	10	16	
			16	18	19	20	22	23	25	18	22					
			26	28	29	34	35	36	37	26	34					
HRV	40	14	2	3	8	13	17	18	21	6	2	3	13	6	2	3
			22	23	25	30	34	39	40	22	23	25				
PPG HRV	86	34	PPG	3	9	10	11	12	16	11	10	16	18	11	10	16
				18	19	20	22	23	25	22						
				26	28	29	34	35	36							
		HRV	2	3	7	8	13	17		2	3	13	2	3	13	
			18	21	22	23	25	28	18	22	23					
			32	35	37	38			25							

FC, feature count; SFC, selected feature count; SFN, selected feature number.

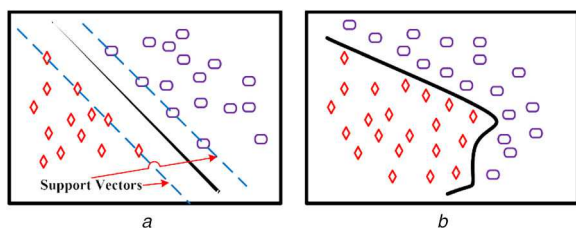


Fig. 8 Separation of classes by (a) Linear, (b) Non-linear lines

Table 6 Network operation parameters

Kernel function		BC	S
Gaussian or RBF kernel	rbf	1	1 True
		2	
Linear kernel	linear	3	
		4	0 False
Polynomial kernel	polynomial	...	
		100	

BC, BoxConstraint; S, standardise; RBF, radial basis function.

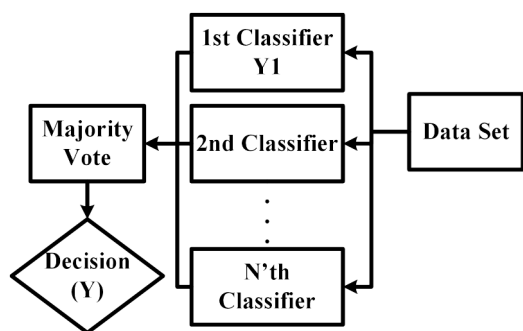


Fig. 9 Ensemble classifier working algorithm

2.5.4 Support vector machines: SVMs is one of the best classification algorithms [32]. Developed in 1995. SVMs can also be used effectively in regression analysis. SVMs tries to separate data sets from each other with linear and non-linear decision boundaries (Fig. 8).

The purpose of SVMs is to be able to distinguish data sets with minimum error [33]. Nearest data to the boundary line separating data sets from each other are the support vectors (Fig. 8).

The network parameters used for the SVMs designed in the study are summarised in Table 6. There are three different kernel functions which are network parameters, a BoxConstraint parameter between 1 and 100, and parameters whether to normalise the data (two different statuses). When these parameters are taken into consideration, in order to classify each data set in the study, a

Table 7 Data distribution in training and test phases in classifiers

Class, %	Sleep/wake detection		
Wake, %	Sleep, %	Total, %	
training (50)	740 (49.97)	741 (50.03)	1481 (100)
test (50)	741 (50.03)	740 (49.97)	1481 (100)
total (100)	1481 (100)	1481 (100)	

total of $3 \times 100 \times 2 = 600$ different network design has been made and the best performing network has been determined and performance evaluation criteria of the network have been calculated.

2.5.5 Ensemble classifier: The ensemble classifier is created by combining several classifiers to produce predictions with high accuracy [34]. The working structure of the ensemble classifier is shown in the flow diagram in Fig. 9. The system is composed of *N* number classifiers. *N* can be single or double. While classifying according to the feature vector, each classifier generates an output value for the first feature vector. The produced output values are counted. The decision of the ensemble classifier is then determined by the majority of votes. If the number of classifiers is doubled, the average value of the classifiers' decision values is rounded off and the decision of the ensemble classifier is determined. This applies to the entire feature vector.

Suppose that in an ensemble classifier with four classifiers, the output values are 1: normal and 2: apnea. If the four classifiers produce 1 1 2 1 outputs, respectively, then the decision value of the ensemble classifier becomes 1 with the majority vote. If the estimated values of the classifiers are 1 1 2 2, respectively, then the arithmetic mean is taken. The average of 1 1 2 2 corresponds to 1.5 and this number corresponds to 2 when rounded. In this case, the output value of the ensemble classifier is set to 2.

The ensemble classifier was prepared in MATLAB environment using four different classifiers which are kNN, MLFFNN, PNN, SVMs [23].

There are 1482 awake epochs and 6953 sleeping books epochs in the database. Since the data are unstable, they are first balanced. For this, 1481 (23.30%) samples were chosen to be close to the data of the other group according to the systematic sampling method of 6953 sleep epochs. This was reduced to 1481 in the other group with 1482.

The data for classification is divided into training and test data sets (Table 7). The data were selected according to the systematic sampling theorem [35].

Two different approaches can be used to determine training and test groups. The first is to create a balanced set of data according to the labels of the data. For this, it should be ensured that the balanced data is collected according to the labels. This study aimed to determine the state of sleep/wake. So, in this study, we have tried to distribute the data of sleep and wake labels balanced for training and testing. The second method is patient-based grouping. In this method, if ten individuals were used in the study, if the education and test data are to be divided into 50–50%, five individuals are included in the education group, and the other five individuals are included in the test group. However, in this case, the data imbalance occurs because there are no balanced sleep/wake records in each patient (Table 1). When the data is grouped in this way, the vast data set is converted to an inefficiently unbalanced data set. This is a significant problem in the literature (unbalanced data classification problem) [7]. Since we think that transforming the data into a problematic data structure will harm the health of the study, we have chosen to distribute the first method, that is, to distribute the sleep/wake records in a balanced manner.

2.6 Used performance criteria

The following criteria were used to evaluate classifiers [12]. Table 8 was given a comparison matrix for accuracy, specificity, and sensitivity: (i) accuracy rates (4), (ii) sensitivity (5), (iii)

specificity (6), (iv) kappa value (Table 9), (v) *F*-measure (7), (vi) receiver operating characteristic (ROC), (vii) area under the ROC curve (AUC), (viii) *k*-fold cross-validation accuracy rate

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \times 100 \quad (4)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (6)$$

$$F = 2 \times \frac{\text{Specificity} \times \text{Sensitivity}}{\text{Specificity} + \text{Sensitivity}} \quad (7)$$

3 Results

In the study, a new approach to artificial intelligence-based sleep alternation, which is the most important diagnostic step of the OSA, has been developed.

For sleep/awake detection, a system based on machine learning was developed with features extracted from PPG and HRV. In addition, in order to increase the performances of the classifiers, the ones that are effective in the 86 items extracted from PPG and HRV were selected by the *F*-score method. *F*-scores were applied to the specifics twice and the effect of the *F*-score method applied at different levels was examined by classifying each step. Table 5 shows the selected PPG and HRV features for sleep/awake detection after the *F*-score method. The total number of features is given in Table 5 in the 'Feature Count' column. In the column titled 'PPG HRV', the extracted features have been combined and used. In total, 46 features extracted from PPG were reduced to 21 features when the first *F*-score was applied. The feature numbers of these 21 features are shown in the column titled 'selected feature numbers'. When *F*-score was applied to PPG for the second time, 21 properties were reduced to 6. The same applies to HRV. A total of 40 features extracted were reduced to 14 in the application of first *F*-score and were reduced to six in the second application. When combined with PPG and HRV, total 86 features were reduced to 34 with the first *F*-score application and 11 with the second application.

After the completion of the feature selection process, the classification process was carried out. Classification procedures were made separately for PPG and HRV and then combined, thus all features were used. Classification results of PPG features are summarised in Table 10, classification results of HRV are in Table 11, classification results of PPG and HRV are in Table 12. For each classifier used in the table, the results were given in detail. When the results of PPG are examined in Table 10, sleep/awake detection success rate for all classifiers was close to 80% while all the features were used. For the ensemble classifier, this value was 85.62%. Likewise, the specificity and sensitivity of the classifiers, in other words, their ability to distinguish sleep and wakefulness, were around 80%. In Table 11, when the classification results of HRV were examined, the success rate of the ensemble classifier was 87.24% while kNN, MLFFNN, and SVMs were around 75%. Although the accuracy of the classifiers decrease when the number of features is reduced, the success rate of the ensemble classifier is 90.01%. This rate was achieved with only 14 features.

Classification procedures were done in order and the sequence is as follows. First, 46 features of PPG were classified without being subjected to any feature selection algorithm and performance parameters were calculated in the related column to measure the performance of the classifier. Then, with the applied first feature selection algorithm, 46 features were reduced to 21 and the same process was repeated. With the second feature selection method, 21 features were reduced to six and reclassification process was performed. The performance criteria of the classifiers were calculated and shown in the table. The number of features obtained when each *F*-score is applied is also given. The *F*-measure and AUC values of these classifiers for all features were around 80%.

Table 8 Comparison matrix for accuracy, specificity and sensitivity

		Estimated	
		Positives (P)	Negatives (N)
factual situation	positives (P) negatives (N)	true positives (TP) false positives (FP)	false negatives (FN) true negatives (TN)

Table 9 Kappa coefficient limit ranges

Kappa coefficient	Explanation
0.81–1.00	very good level of integration
0.61–0.80	good level of integration
0.41–0.60	medium level of integration
0.21–0.40	low level of integration
0.00–0.20	weak level of integration
<0.00	very weak level of integration

Each classifier in the 'network parameters' tab was calculated based on the classifier's network parameter.

ROC curves of the classifiers were also prepared and shown in Figs. 10–18. The ROC curve can be interpreted as follows. In ROC curves, if the curve is closer to the left axis it is apnea, if the curve is closer to the upper axis, it can be said that the system is able to diagnose the control group better.

The general evaluation table for classification results is shown in Table 13, and the graph is shown in Fig. 19. The best performance for sleep/awake detection according to Table 12 is obtained with the ensemble classifier. Only HRV can be used alone for this processor PPG and HRV can be used together. The combined use of PPG and HRV both improved the classification performance and reduced the number of features, which is an advantage. The best performance for sleep/awake detection was provided by the ensemble classifier using 11 features of PPG (4) and HRV (7). This process can also be done with 14 HRV features.

However, both PPG and HRV features (11) can be used because both performances are high and featured workload is low. It is possible to develop a physical system according to the results obtained.

The analysis results in Fig. 20 clearly show the positive effect of the *F*-score feature selection algorithm on performance. Here, the *F*-score increased accuracy in some cases and decreased in some cases.

4 Discussion

When a new method for diagnosing any disease is developed, it is compared with reference methods. In total, 80% similarity for the comparison is considered to be evidence that the new method is feasible [36–38]. Most of the performance criterion values of the method developed in this study were over 80%. The reference to the study was the diagnosis of specialist physician. This result is expected and achievable in the literature [10–12, 39]. The most obvious advantage of this work was that machine learning techniques were used in the Ensemble classifier model. On this count, the performance of the system was increased. In this study, the comparison of the proposed method with the reference method for sleep/awake detection was shown in different tables for PPG and HRV. A total of 86 features were extracted in the study. However, considering the difficulty of extracting such features in real time systems, it was tried to develop the system by reducing the features. Ultimately, the number of features was reduced from 86 to 6. The reason for having a large number of features was to be able to specify every detail to cover the signal. Reducing the number of features afterwards allowed us to obtain useful features. This study is one step ahead of the literature in terms of feature count [40–42]. It can be said that this is the perfect performance for a system that can be practically implemented.

The specificity and sensitivity of the classifiers, in other words, their ability to distinguish sleep and wakefulness, were around 80%. This rate is over 80% in the ensemble classifier. As both

Table 10 Classifier results of PPG features for sleep/awake detection

kNN algorithm									
NP	$k = 1$, DF = 'mahalanobis'			$k = 6$, DF = 'mahalanobis'			$k = 6$, DF = 'mahalanobis'		
Class	NFSA			FSA			FSA		
	Feature count = 46			Feature count = 21			Feature count = 6		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.76	0.84	79.81	0.81	0.81	80.96	0.82	0.77	79.61
sleep	0.84	0.76		0.81	0.81		0.77	0.82	
AUC		0.80			0.81			0.80	
Kappa		0.60			0.62			0.59	
FM		0.80			0.81			0.80	
kF		73.06			78.33			73.40	

MLFFNN									
NP	NN = 27, TA = 'trainIm'			NN = 46, TA = 'trainIm'			NN = 69, TA = 'trainIm'		
Class	NFSA			FSA			FSA		
	Feature count = 46			Feature count = 21			Feature count = 6		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.75	0.83	79.34	0.73	0.86	79.27	0.71	0.84	77.52
sleep	0.83	0.75		0.86	0.73		0.84	0.71	
AUC		0.79			0.79			0.78	
Kappa		0.59			0.59			0.55	
FM		0.79			0.79			0.77	
kF		—			—			—	

PNN									
NP	Spread = 0.3890			Spread = 0.1570			Spread = 0.0580		
Class	NFSA			FSA			FSA		
	Feature count = 46			Feature count = 21			Feature count = 6		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.63	0.85	74.07	0.76	0.78	77.11	0.69	0.85	76.71
sleep	0.85	0.63		0.78	0.76		0.85	0.69	
AUC		0.74			0.77			0.77	
Kappa		0.48			0.54			0.53	
FM		0.72			0.77			0.76	
kF		—			—			—	

SVMs									
NP	Kernel = 'rbf', BC = 32			Kernel = 'rbf', BC = 46			Kernel = 'rbf', BC = 2		
Class	NFSA			FSA			FSA		
	Feature count = 46			Feature count = 21			Feature count = 6		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.79	0.77	78.33	0.81	0.81	81.04	0.81	0.78	79.14
sleep	0.77	0.79		0.81	0.81		0.78	0.81	
AUC		0.78			0.81			0.79	
Kappa		0.57			0.62			0.58	
FM		0.78			0.81			0.79	
kF		77.52			78.46			75.83	

specificity and sensitivity increase, the reliability of the method is considered to be increased [36–38]. The specificity and sensitivity values of 0.97 and 0.85 of the developed system appear to be above the expected value when compared with the literature [5, 41, 42].

In addition to the accuracy of the classifiers, the F -measure and AUC values of these classifiers for all features are around 80%. A different evaluation performance criterion is the ROC curve. The reliability of the system is reinforced by these parameters. ROC curves for systems developed for sleep/awake detection are shown

in Figs. 10–18. Nine different ROC curves are shown for the nine classified data sets.

When curves are evaluated, the ideal ROC curve found in the graph is referenced. The closest classifier to this curve is the best. ROC curves of developed classifiers are fairly close to ideal. In addition to all these good aspects of the system, the Kappa value of the classifier is quite low. In this regard, the system can be further improved. For the development of the system different features can be extracted from PPG and HRV, which can represent the sleep stages. In addition, the database used can be expanded. PPG is a

Classifier ensemble									
Class	NFSA			FSA			FSA		
	Feature count = 46			Feature count = 21			Feature count = 6		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.77	0.93	85.62	0.80	0.94	86.90	0.78	0.97	87.17
sleep	0.93	0.77		0.94	0.80		0.97	0.78	
AUC		0.85			0.87			0.88	
Kappa		0.71			0.74			0.75	
FM		0.84			0.86			0.87	
kF		—			—			—	

NP, network parameters; FM, F-measure; kF, k(10)-fold (%); DF, distance function; NFSA, no feature selection applied; FSA, feature selection applied; Sen, sensitivity; Spe, specificity; Acc, accuracy; NN, number of neurons; TA, training algorithm; BC, BoxConstraint.

Table 11 Classifier results of HRV characteristics for sleep/awake detection

kNN algorithm									
Class	$k = 10$, DF = 'cityblock'			$k = 6$, DF = 'cosine'			$k = 3$, DF = 'cityblock'		
	NFSA			FSA			FSA		
	Feature count = 40			Feature count = 14			Feature count = 6		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.75	0.74	74.54	0.72	0.75	73.87	0.72	0.73	72.65
sleep	0.74	0.75		0.75	0.72		0.73	0.72	
AUC		0.75			0.74			0.73	
Kappa		0.49			0.48			0.45	
FM		0.75			0.74			0.73	
kF		68.94			70.83			69.07	

MLFFNN									
Class	NN = 47, TA = 'trainlm'			NN = 38, TA = 'trainlm'			NN = 51, TA = 'trainlm'		
	NFSA			FSA			FSA		
	Feature count = 40			Feature count = 14			Feature count = 6		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.70	0.83	76.30	0.68	0.78	72.72	0.63	0.80	71.24
asleep	0.83	0.70		0.78	0.68		0.80	0.63	
AUC		0.76			0.73			0.71	
Kappa		0.53			0.45			0.42	
FM		0.76			0.72			0.70	
kF		—			—			—	

PNN									
Class	Spread = 0.2460			Spread = 0.0980			Spread = 0.0290		
	NFSA			FSA			FSA		
	Feature count = 40			Feature count = 14			Feature count = 6		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.65	0.78	71.30	0.69	0.78	73.19	0.72	0.73	72.59
sleep	0.78	0.65		0.78	0.69		0.73	0.72	
AUC		0.71			0.73			0.73	
Kappa		0.43			0.46			0.45	
FM		0.71			0.73			0.73	
kF		—			—			—	

new-used signal for sleep/awake detection. In this respect, the studies in which PPG is used for sleep/awake detection in the literature are new and quite scarce [11, 12]. HRV sleep/awake detection, which can be derived from ECG, is frequently used [10, 12]. However, deriving from ECG is more difficult than PPG [43]. In the studies on the comparison of the HRV signal derived from the ECG and PPG, there is a 0.99 correlation between the HRV signals derived from both signals [43]. In other words, it is possible to perform the same operation (such as sleep staging) with HRV

derived from both signals. Also, since the derived HRV signals are the same, the performance values will be the same. Therefore, the same high performance would be achieved if this work was done with ECG.

In addition, correlations of HRV signals that were extracted from PPG and ECG were 0.99 [43]. HRV used in this study is derived from PPG. This is not a disadvantage, but rather an advantage. As stated in the literature, PPG has much extra information compared to ECG, such as respiration rate and value of

SVMs									
NP	Kernel = 'rbf', BC = 10			Kernel = 'rbf', BC = 4			Kernel = 'rbf', BC = 82		
Class	NFSA			FSA			FSA		
	Feature count = 40			Feature count = 14			Feature count = 6		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.74	0.76	75.02	0.73	0.74	73.60	0.77	0.71	74.27
sleep	0.76	0.74		0.74	0.73		0.71	0.77	
AUC		0.75			0.74			0.74	
Kappa		0.50			0.47			0.49	
FM		0.75			0.74			0.74	
kF		71.84			70.56			71.37	

Ensemble classifier									
Class	NFSA			FSA			FSA		
	Feature count = 40			Feature count = 14			Feature count = 6		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.78	0.97	87.24	0.82	0.97	90.01	0.79	0.95	87.24
sleep	0.97	0.78		0.97	0.82		0.95	0.79	
AUC		0.87			0.90			0.87	
Kappa		0.75			0.80			0.74	
FM		0.86			0.89			0.86	
kF		—			—			—	

NP, network parameters; FM, *F*-measure; kF, k(10)-fold (%); DF, distance function; NFSA, no feature selection applied; FSA, feature selection applied; Sen, sensitivity; Spe, specificity; Acc, accuracy; NN, number of neurons, TA, training algorithm; BC, BoxConstraint.

Table 12 Classifier results of PPG and HRV for sleep/awake detection

kNN algorithm									
NP	<i>k</i> = 7, DF = 'cityblock'			<i>k</i> = 5, DF = 'mahalanobis'			<i>k</i> = 10, DF = 'mahalanobis'		
Class	NFSA			FSA			FSA		
	Feature count = 86			Feature count = 34			Feature count = 11		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.75	0.84	79.68	0.77	0.84	80.96	0.81	0.80	80.49
sleep	0.84	0.75		0.84	0.77		0.80	0.81	
AUC		0.80			0.81			0.80	
Kappa		0.59			0.62			0.61	
FM		0.79			0.81			0.80	
kF		74.88			77.65			75.83	

MLFFNN									
NP	NN = 23, TA = 'trainlm'			NN = 57, TA = 'trainlm'			NN = 23, TA = 'trainlm'		
Class	NFSA			FSA			FSA		
	Feature count = 86			Feature count = 34			Feature count = 11		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.79	0.84	81.57	0.78	0.84	81.03	0.73	0.85	79.07
sleep	0.84	0.79		0.84	0.78		0.85	0.73	
AUC		0.82			0.81			0.79	
Kappa		0.63			0.62			0.58	
FM		0.82			0.81			0.79	
kF		—			—			—	

oxygen saturation in the blood [43]. Therefore, this study is a step ahead in terms of creating convenience for HRV derivation [10].

According to the results obtained in this study, it was concluded that PPG and HRV derived from this signal can be used for sleep/wake detection and produce meaningful results. The easy acquisition of the PPG signal and the derivation of the HRV from the PPG signal opens up the possibility of performing respiratory scoring with a single signal. In systems that can operate in real time, easy measurement and easy processing of the signal will

increase the practicality of the systems. In OSA diagnosis, at least 16 channel signals are needed for sleep staging. The use of PPG signals instead of these channel signals will reduce the workload.

The sleep/awake detection process can be performed with PPG and HRV features at an accuracy of 71–91%. The system can be realised with 11 PPG and HRV specification and accuracy of 91.09%. This success rate is also sufficient for the system to work [36–38]. When all these values are taken into consideration, it is possible that a practical sleep-staging system can be realised.

PNN									
NP	Spread = 0.3560			Spread = 0.1610			Spread = 0.0950		
Class	NFSA			FSA			FSA		
	Feature count = 86			Feature count = 34			Feature count = 11		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.72	0.84	77.65	0.81	0.78	79.61	0.72	0.85	78.19
sleep	0.84	0.72		0.78	0.81		0.85	0.72	
AUC		0.78			0.80			0.78	
Kappa		0.55			0.59			0.56	
FM		0.77			0.80			0.78	
kF		—			—			—	

SMVs									
NP	Kernel = 'polynomial', BC = 1			Kernel = 'rbf', BC = 2			Kernel = 'rbf', BC = 2		
Class	NFSA			FSA			FSA		
	Feature count = 86			Feature count = 34			Feature count = 11		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.76	0.85	80.49	0.80	0.81	80.42	0.85	0.79	82.11
sleep	0.85	0.76		0.81	0.80		0.79	0.85	
AUC		0.80			0.80			0.82	
Kappa		0.61			0.61			0.64	
FM		0.80			0.80			0.82	
kF		78.12			77.79			78.39	

Ensemble classifier									
Class	NFSA			FSA			FSA		
	Feature count = 86			Feature count = 34			Feature count = 11		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %	Sen	Spe	Acc, %
awake	0.85	0.96	91.09	0.84	0.92	88.39	0.85	0.97	91.09
sleep	0.96	0.85		0.92	0.84		0.97	0.85	
AUC		0.91			0.88			0.91	
Kappa		0.82			0.77			0.82	
FM		0.90			0.88			0.91	
kF		—			—			—	

NP, network parameters; FM, *F*-measure; kF, *k*(10)-fold (%); DF, distance function; NFSA, no feature selection applied; FSA, feature selection applied; Sen, sensitivity; Spe, specificity; Acc, accuracy; NN, number of neurons; TA, training algorithm; BC, BoxConstraint.

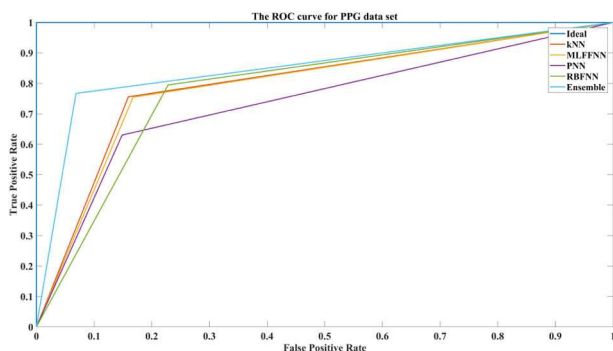


Fig. 10 ROC curve for all PPG features (46)

Graphic summary representation of numerical values was given in Table 13 and Fig. 19.

F-score can affect performance in two ways. In the first case, the success rate increases while the number of features decreases. In the second case, the success rate decreases while the number of features decreases. System performance may improve in both cases. The reduced success rate is acceptable, if not an excessive reduction, according to the problem. In both cases, the system workload will be reduced as the number of features is reduced. In our problem, the success rate with minimum features is quite high.

Reducing the number of features will reduce the processing load. If the processing load is reduced, the small decrease in success rate is acceptable for our problem. The acceptable success rate for this problem is ~70–80%.

The ensemble classifier provided superior performance in classifying each group of data. This is because the other classifier compensates for the mistake made by a classifier when performing individual classification operations. In this way, the system's power has been increased by combining individual performances. Also, the distributions of the data were designed so that there is no difference between the groups. Normal and smooth distributions had a positive effect on the system in order to work better [8].

5 Conclusion

This study suggests that the PPG signal can be used to diagnose OSA by machine learning and signal processing techniques. In the literature, quite different signals and combinations are used for the diagnosis of OSA. However, measuring the signal to be used with easy and non-invasive methods will reduce the discomfort experienced by the patient.

The advantages of the study can be explained as follows. The study can be developed from different angles. The PPG and HRV feature extracted from this study can be used when any OSA diagnostic system is being built. In this way, the workload required

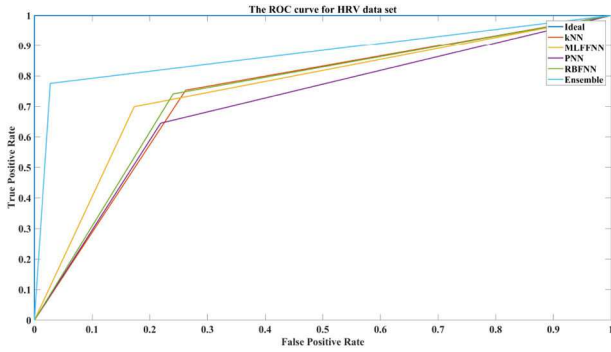


Fig. 11 ROC curve for all HRV features (40)

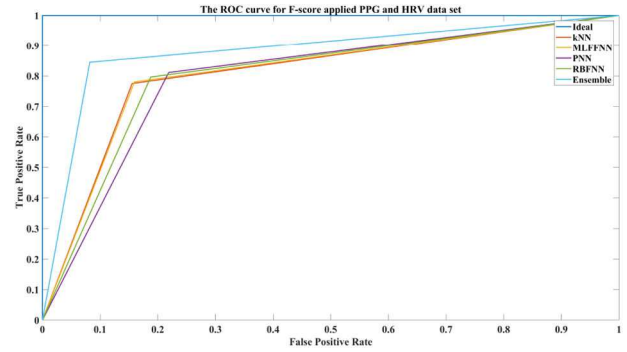


Fig. 15 ROC curve for 34 PPG (18) and HRV (16) features

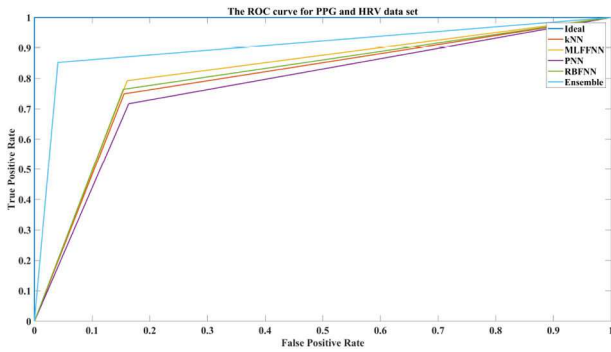


Fig. 12 ROC curve for all PPG and HRV features (86)

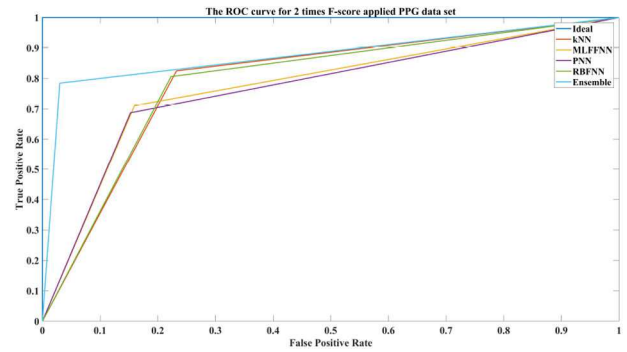


Fig. 16 ROC curve for 6 PPG features

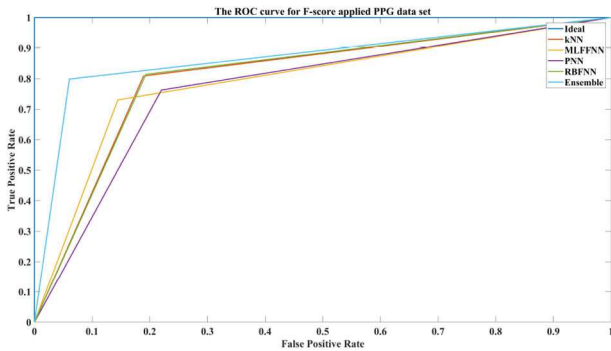


Fig. 13 ROC curve for 21 PPG features

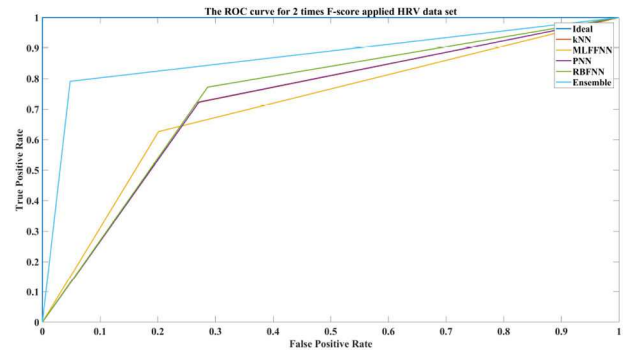


Fig. 17 ROC curve for 6 HRV features

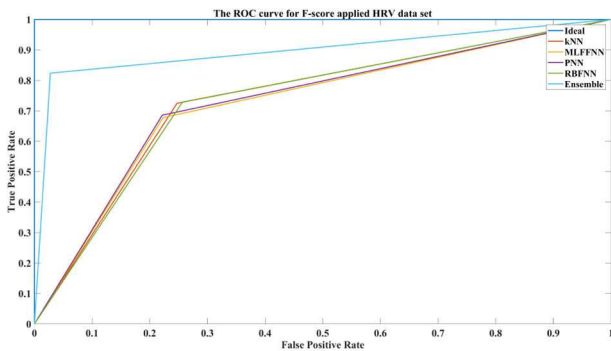


Fig. 14 ROC curve for 14 HRV features

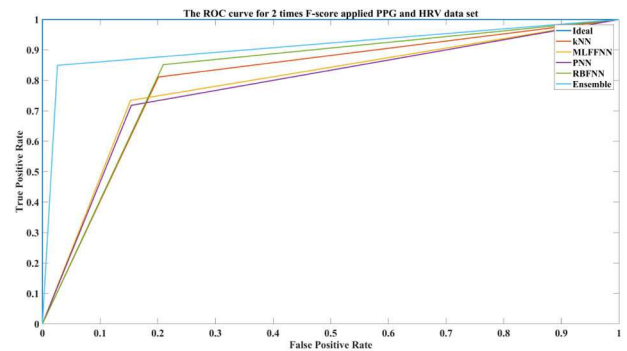


Fig. 18 ROC curve for 11 PPG (4) and HRV (7) features

for OSA diagnosis can be reduced by designing a system capable of real-time analysis. In addition, the system can be easily used by the patient without the need for technical personnel. The availability of the system at home is a different advantage. Carrying out the diagnosis in a shorter amount of time means a quick start to treatment. Thus, the ailments given by the OSA to the human body can be avoided in time. Despite these advantages, the disadvantage of operation is that the signal processing load is high. Therefore, checking the analysis can improve system quality.

6 Future work

This study can be extended further by repeating different signals, properties, feature selection algorithms and machine learning algorithms. In the field of health, it may be advantageous to work with large data.

Future studies can be listed as follows.

- More efficient signal processing and feature extraction technique design
- Extend the work with different machine learning algorithms
- Developing fewer features and higher performance systems with different feature selection algorithms

Table 13 Summary classifier results of the PPG and the HRV features

Ensemble classifier Signal	PPG and HRV			HRV		
	Feature count = 11			Feature count = 14		
	Sen	Spe	Acc, %	Sen	Spe	Acc, %
Apnea	0.85	0.97	91.09	0.82	0.97	90.01
Kontrol	0.97	0.85		0.97	0.82	
AUC		0.91			0.90	
Kappa		0.82			0.80	
F-measure		0.91			0.89	
k(10)-fold, %		—			—	

Sen, sensitivity; Spe, specificity; Acc, accuracy.

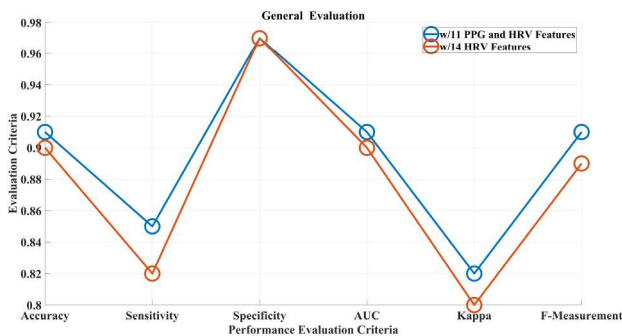


Fig. 19 General evaluation graph for classification results

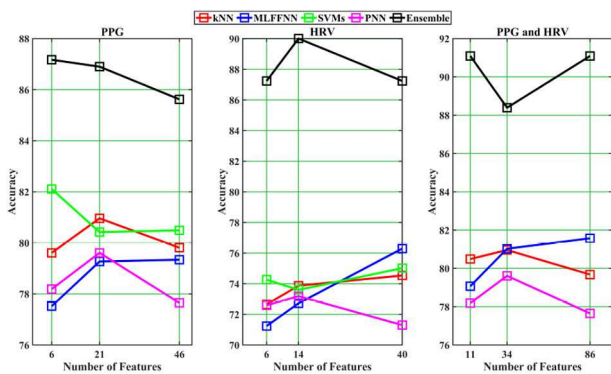


Fig. 20 Performance by the number of selected features

7 Acknowledgments

This research was supported by The Scientific and Technical Research Council of Turkey (TUBITAK) through The Research Support Programs Directorate (ARDEB) with project number of 115E657, and project name of ‘A New System for Diagnosing Obstructive Sleep Apnea Syndrome by Automatic Sleep Staging Using Photoplethysmography (PPG) Signals and Breathing Scoring’ and by The Coordination Unit of Scientific Research Projects of Sakarya University.

The ethics committee report numbered 16214662/050.01.04/70 from Sakarya University Deanship of Faculty of Medicine, and the data use permission numbered 94556916/904/151.5815 from T.C. Ministry of Health Turkey Public Hospitals Authority Sakarya Province General Secretariat of Association of Public Hospitals were received to perform the study.

8 References

[1] Berry, R.B., Budhiraja, R., Gottlieb, D.J., et al.: ‘Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine’, *J. Clin. Sleep Med.*, 2012, **8**, (5), pp. 597–619

[2] Song, C., Liu, K., Zhang, X., et al.: ‘An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals’, *IEEE Trans. Biomed. Eng.*, 2016, **63**, (7), pp. 1532–1542

[3] Bruyneel, M., Ninane, V.: ‘Unattended home-based polysomnography for sleep disordered breathing: current concepts and perspectives’, *Sleep Med. Rev.*, 2014, **18**, (4), pp. 341–347

[4] Hassan, A.R., Bhuiyan, M.I.H.: ‘A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features’, *J. Neurosci. Methods*, 2016, **271**, pp. 107–118

[5] Hassan, A.R., Bhuiyan, M.I.H.: ‘An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting’, *Neurocomputing*, 2017, **219**, pp. 76–87

[6] Ebrahimi, F., Setarehdan, S.-K., Nazeran, H.: ‘Automatic sleep staging by simultaneous analysis of ECG and respiratory signals in long epochs’, *Biomed. Signal Process. Control*, 2015, **18**, pp. 69–79

[7] Duan, L., Xie, M., Bai, T., et al.: ‘A new support vector data description method for machinery fault diagnosis with unbalanced datasets’, *Expert Syst. Appl.*, 2016, **64**, pp. 239–246

[8] Liu, Z., Sun, J., Zhang, Y., et al.: ‘Sleep staging from the EEG signal using multi-domain feature extraction’, *Biomed. Signal Process. Control*, 2016, **30**, pp. 86–97

[9] Tripathy, R.K., Rajendra Acharya, U.: ‘Use of features from RR-time series and EEG signals for automated classification of sleep stages in deep neural network framework’, *Biocybern. Biomed. Eng.*, 2018, **38**, (4), pp. 890–902

[10] Hayet, W., Slim, Y.: ‘Sleep-wake stages classification based on heart rate variability’. 2012 5th Int. Conf. on BioMedical Engineering and Informatics, Chongqing, China, 2012, pp. 996–999

[11] Dehkordi, P., Garde, A., Karlen, W., et al.: ‘Sleep stage classification in children using photoplethysmogram pulse rate variability’. Computing in Cardiology Conf. (CinC), Massachusetts, USA, 2014, pp. 297–300

[12] Uçar, M.K., Bozkurt, M.R., Bilgin, C., et al.: ‘Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques’, *Neural Comput. Appl.*, 2018, **29**, pp. 1–16

[13] Tripathy, R.K.: ‘Application of intrinsic band function technique for automated detection of sleep apnea using HRV and EDR signals’, *Biocybern. Biomed. Eng.*, 2018, **38**, (1), pp. 136–144

[14] Uçar, M.K., Bozkurt, M.R., Bilgin, C., et al.: ‘Automatic detection of respiratory arrests in OSA patients using PPG and machine learning techniques’, *Neural Comput. Appl.*, 2017, **28**, (10), pp. 2931–2945

[15] Anderer, P., Moreau, A., Woertz, M., et al.: ‘Computer-assisted sleep classification according to the standard of the American Academy of Sleep Medicine: validation study of the AASM version of the Somnolyzer 24 × 7’, *Neuropsychobiology*, 2010, **62**, (4), pp. 250–264

[16] Fraiwan, L., Lweesy, K., Khasawneh, N., et al.: ‘Classification of sleep stages using multi-wavelet time frequency entropy and LDA’, *Methods Inf. Med.*, 2010, **49**, (3), pp. 230–237

[17] Berthomier, C., Drouot, X., Herman-Stoica, M., et al.: ‘Automatic analysis of single-channel sleep EEG: validation in healthy individuals’, *Sleep*, 2007, **30**, (11), pp. 1587–1595

[18] Schaltenbrand, N., Lengelle, R., Toussaint, M., et al.: ‘Sleep stage scoring using the neural network model: comparison between visual and automatic analysis in normal subjects and patients’, *Sleep*, 1996, **19**, (1), pp. 26–35

[19] Wang, Y., Loparo, K.A., Kelly, M.R., et al.: ‘Evaluation of an automated single-channel sleep staging algorithm’, *Nat. Sci. Sleep*, 2015, **7**, pp. 101–111

[20] Ahuja, N.D., Agarwal, A.K., Mahajan, N.M., et al.: ‘GSR and HRV: its application in clinical diagnosis’. 16th IEEE Symp. Computer-Based Medical Systems, 2003. Proc., New York, USA, 2003, pp. 279–283

[21] Elgendi, M.: ‘On the analysis of fingertip photoplethysmogram signals’, *Curr. Cardiol. Rev.*, 2012, **8**, (1), pp. 14–25

[22] Alian, A.A., Shelley, K.H.: ‘Photoplethysmography’, *Best Pract. Res. Clin. Anaesthesiol.*, 2014, **28**, (4), pp. 395–406

[23] Wallisch, P., Lusignan, M.E., Benayoun, M.D., et al.: ‘MATLAB for neuroscientists’ (Elsevier, USA, 2014)

[24] Shi, P., Zhu, Y., Allen, J., et al.: ‘Analysis of pulse rate variability derived from photoplethysmography with the combination of lagged Poincaré plots and spectral characteristics’, *Med. Eng. Phys.*, 2009, **31**, (7), pp. 866–871

[25] Polat, K., Güneş, S.: ‘A new feature selection method on classification of medical datasets: kernel F-score feature selection’, *Expert Syst. Appl.*, 2009, **36**, (7), pp. 10367–10373

[26] Şahan, S., Polat, K., Kodaz, H., et al.: ‘A new hybrid method based on fuzzy-artificial immune system and k-NN algorithm for breast cancer diagnosis’, *Comput. Biol. Med.*, 2007, **37**, (3), pp. 415–423

[27] Khan, M., Ding, Q., Perrizo, W.: ‘k-nearest neighbor classification on spatial data streams using P-trees’, in Ming, S., Chen/Philip, S., YuBing, L. (Eds.):

- 'Advances in knowledge discovery and data mining' (Springer: Berlin, Heidelberg, 2002), pp. 517–528
- [28] Farajzadeh, J., Fakheri Fard, A., Lotfi, S.: 'Modeling of monthly rainfall and runoff of Urmia lake basin using "feed-forward neural network" and "time series analysis" model', *Water Resour. Ind.*, 2014, **7–8**, pp. 38–48
- [29] Zounemat-kermani, M., Kisi, O., Rajaei, T.: 'Performance of radial basis and LM-feed forward artificial neural networks for predicting daily watershed runoff', *Appl. Soft Comput.*, 2013, **13**, (12), pp. 4633–4644
- [30] Khamis, A., Shareef, H., Mohamed, A., *et al.*: 'Islanding detection in a distributed generation integrated power system using phase space technique and probabilistic neural network', *Neurocomputing*, 2015, **148**, pp. 587–599
- [31] Wasserman, P.D.: '*Advanced methods in neural computing*' (Van Nostrand Reinhold, USA, 1993)
- [32] Cortes, C., Vapnik, V.: 'Support-vector networks', *Mach. Learn.*, 1995, **20**, pp. 273–297
- [33] Mandhala, V.N., Sujatha, V., Devi, B.R.: 'Scene classification using support vector machines'. 2014 IEEE Int. Conf. on Advanced Communications, Control and Computing Technologies, Ramanathapuram, India, 2014, pp. 1807–1810
- [34] Rokach, L., Schlar, A., Itach, E.: 'Ensemble methods for multi-label classification', *Expert Syst. Appl.*, 2014, **41**, (16), pp. 7507–7523
- [35] Ramachandran, K.M., Tsokos, C.P.: '*Mathematical statistics with applications in R*' (Elsevier, USA, 2015)
- [36] Alpar, R.: '*Applied statistic and validation – reliability*' (Detay Publishing, Turkey, 2010)
- [37] Van Stralen, K.J., Stel, V.S., Reitsma, J.B., *et al.*: 'Diagnostic methods I: sensitivity, specificity, and other measures of accuracy', *Int. Soc. Nephrol.*, 2009, **75**, (12), pp. 1257–1263
- [38] Šimundić, A.-M.: 'Measures of diagnostic accuracy: basic definitions', *EJIFCC*, 2009, **19**, (4), pp. 203–211
- [39] Sharma, H., Sharma, K.K.: 'An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions', *Comput. Biol. Med.*, 2016, **77**, pp. 116–124
- [40] Ebrahimi, F., Setarehdan, S.K., Ayala-Moyeda, J., *et al.*: 'Automatic sleep staging using empirical mode decomposition, discrete wavelet transform, time-domain, and nonlinear dynamics features of heart rate variability signals', *Comput. Methods Programs Biomed.*, 2013, **112**, (1), pp. 47–57
- [41] Hassan, A.R., Haque, M.A.: 'An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting', *Neurocomputing*, 2017, **235**, pp. 122–130
- [42] Hassan, A.R., Bhuiyan, M.I.H.: 'Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting', *Comput. Methods Programs Biomed.*, 2017, **140**, pp. 201–210
- [43] Lu, G., Yang, F., Taylor, J.A., *et al.*: 'A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects', *J. Med. Eng. Technol.*, 2009, **33**, (8), pp. 634–641