

ONLINE OBSTRUCTIVE SLEEP APNEA DETECTION BASED ON HYBRID MACHINE LEARNING AND CLASSIFIER COMBINATION FOR HOME-BASED APPLICATIONS

Hosna Ghandeharioun

Department of Electrical and Biomedical Engineering,
Khorasan Institute of Higher Education, Mashhad, Iran

ABSTRACT

Automatic detection of obstructive sleep apnea (OSA) is in great demand. OSA is one of the most prevalent diseases of the current century and established comorbidity to Covid-19. OSA is characterized by complete or relative breathing pauses during sleep. According to medical observations, if OSA remained unrecognized and un-treated, it may lead to physical and mental complications. The gold standard of scoring OSA severity is the time-consuming and expensive method of polysomnography (PSG). The idea of online home-based surveillance of OSA is welcome. It serves as an effective way for spurred detection and reference of patients to sleep clinics. In addition, it can perform automatic control of the therapeutic/assistive devices. In this paper, several configurations for online OSA detection are proposed. The best configuration uses both ECG and SpO2 signals for feature extraction and MI analysis for feature reduction. Various methods of supervised machine learning are exploited for classification. Finally, to reach the best result, the most successful classifiers in sensitivity and specificity are combined in groups of three members with four different combination methods. The proposed method has advantages like limited use of biological signals, automatic detection, online working scheme, and uniform and acceptable performance (over 85%) in all the employed databases. These advantages have not been integrated in previous published methods.

KEYWORDS

Obstructive Sleep Apnea, Supervised Machine Learning, Feature Reduction, Classifier Combination, Biomedical Signal Processing.

1. INTRODUCTION

Obstructive sleep apnea (OSA) is the most prevalent sleep-related breathing disorder worldwide [1]. It has also established as a comorbidity to Covid-19 [2]. Intermittent episodes of airway subsidence during sleep characterizes OSA [3]. If OSA remains undetected and untreated, the resultant abrupt changes in sympathetic neural activity may cause severe cardiovascular side-effects [4], type 2 diabetes [5], impaired cognition, and psychiatric symptoms [6]. Hence the detection and immediate treatment of OSA is essential. The diagnosis of OSA requires the joint evaluation of related clinical features and the visible demonstrations of abnormal breathing during sleep. [7]. The gold standard for the detection of abnormal breathing during sleep is overnight polysomnography (PSG). The PSG-driven apnea-hypopnea index (AHI) characterizes the OSA severity [8, 9]. AHI derivation is currently performed visually according to the American Association of Sleep Medicine (AASM) guidelines [8]. This time-consuming and

expensive process imposes a heavy burden on the public health section [10]. Therefore, many automatic methods for pre-clinic detection and scoring of OSA have been developed in the literature [11-27, 31, 32, 34, 35]. These methods use analysis of a variety of biological signals and machine learning techniques. In some studies, electroencephalogram (EEG) is used for feature extraction based on occurred discrepancies between the right and left hemispheres [11] or tracking non-linear behavior of EEG due to fluctuations in sleep depth [12, 29, 30]. Single-channel ECG or combination of ECG and saturated oxygen level of the blood in peripheral veins (SpO₂) is also suggested in several studies due to easy and unobtrusive signal acquisition [13, 14, 16-19, 23].

In the most recent studies, OSA detection is accomplished based on ECG and the newly widespread deep learning techniques. In deep learning solutions, the feature extraction/selection is generally embedded in the learning algorithm, and no separate step is needed [32]. This advantage reduces the computational load. However, for deep learning training, high-performance computers are required [33], and the methodologies do not suit home-based and portable applications where the processing ability and data storage capacity are limited. Apnea is detected based on nasal pressure signals with the help of convolutional neural networks (CNN) in [34]. Several supervised machine learning methods are tested for OSA detection with a single channel ECG signal in [35]. The achieved results are promising, yet in a small database and with slightly less accuracy than our suggested strategy.

In this study, several configurations for online detection of OSA are suggested. Employing a limited number of biological signals, automatic and real-time detection, and uniform acceptable performance over several databases are the merits of our proposed method. To the knowledge of the author, these advantages are accumulated in none of the previous studies together.

2. MATERIAL AND METHOD

Automatic detection of respiratory events based on supervised machine learning is generally divided into several steps [28]. In the first step, the training set is made from signal records labeled as apnoeic and normal (by an expert clinician). In the second step, feature extraction is performed for each signal. The extracted features can be reduced to improve the performance of the next step. Finally, the last step is the classification of the test records. We will go through each step of our work in detail.

We conducted this study based on three databases. The first two databases are public and can be reached by anyone: St. Vincent, University College Dublin (UCD) database [36], eight subjects of Apnea-ECG database [37] whose data include more signals than one ECG channel. The third database is exclusively at our disposal. This database includes clinical records of the sleep laboratory of Ibn-e-Sina Hospital, Mashhad, Iran, from July 2012 to May 2014. The study was approved by the ethics committee overseeing the research proposal (permission no.92/620792, date 2014/03/07). We were allowed to use clinical data only, with no deviation from AASM protocol. The PSG (model: Alice LE, part no. 1002387, Philips Respironics) recordings were conducted in baseline montage with 16 channels on the 158 referred patients. Out of all participants, 134 subjects were diagnosed with OSA, and 24 healthy according to the International Classification of Sleep Disorders II (ICSD-II) [8]. We ascertained sleep apneas as ≥ 10 s of airflow pauses and hypopnea as a $\geq 3\%$ of oxygen desaturation/or arousal preceded by a 50% decrement in the amplitude of baseline airflow. From now on, we refer to this database as “the exclusive database”. Figure 1 shows a 1-minute frame of polysomnographic records of our exclusive database.

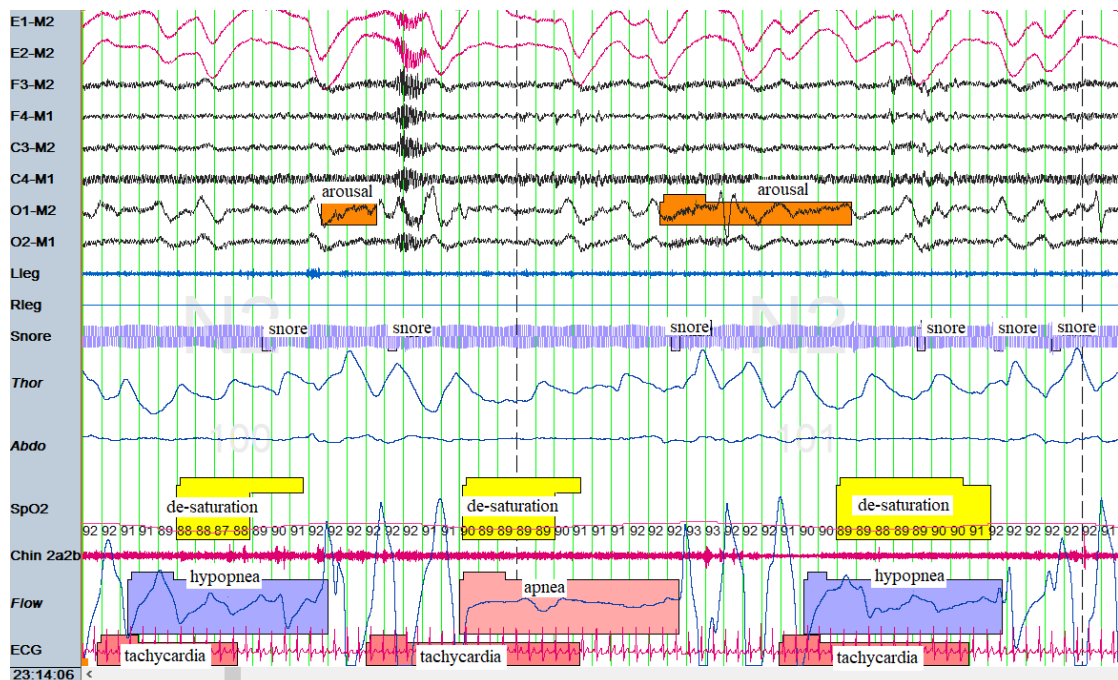


Figure 1. 1-minute frame of the polysomnographic records of a subject with severe OSA from the exclusive database

The three signals of EEG, SpO₂, and air pressure/flow have a central role in the clinical definition of apnea. We refer to them as “the main signals”. Other biological signals (such as ECG, voice, and actigraphy) have a supplementary role in the detection of OSA. We refer to this group as “the auxiliary signals”. Relying on the main signals for an OSA detection system is the first choice; however, the developed system must be more concise than PSG and perform a pre-clinic screening. Placing EEG electrodes on the scalp during sleep and pressure/flow sensors is rather obtrusive; besides, preparations and installation of electrodes and sensors are not straightforward for an ordinary user. For EEG acquisition and conditioning, a relatively expensive system is needed. The repeatability of the observed effects of OSA on EEG compared to SpO₂ signal is also on debate [38, 39]. That is why generally EEG and air pressure/flow signals are excluded.

Among the auxiliary signals, ECG is gained more attention in the OSA detection methods. The effects of the apneas on ECG signal are well understood [4]. The ECG electrodes are installed simpler than EEG and less obtrusive than those of air pressure/flow signal. The apparent effect of respiratory events on ECG is called Cyclical Variation of Heart Rate (CVHR) [4]. The challenge of ECG-based detection systems is their lower specificity since their modulating factor is not a respiratory event only. The presence of cardiovascular problems can also have considerable effects on ECG. In the absence of OSA, these effects can increase the false positive detection rate. In practice, the number of false-negative detections also increases, and the sensitivity of the OSA detection method drops. A decrease in sensitivity is because the database usually includes subjects with OSA whose problem has been un-diagnosed for years, and lack of treatment has led to cardiovascular complexities for them [4]. Up to 90% of subjects affected by OSA are not aware of their problem and have not been treated yet [1].

More successful results are reported for SpO₂-based detection methods compared to other single-channel detection systems. They have reasonable specificity and sensitivity, they can be

performed in real-time, and they have non-obtrusive sensors; additionally, some of them are realized in smartphones and can serve as useful home-based systems [27, 40].

In this study, we consider PPG (and SpO₂) from “the main signals”, and ECG from “the auxiliary signals”. Parallel use of these signals, covers their deficiencies and increases the overall accuracy, sensitivity, and specificity of the detection system [16]. The OSA detection based on ECG and SpO₂ is more popular than other multi-channel detection systems due to simple sensor installation and powerful representation of respiratory events [13, 14, 16-19, 23].

2.1. Pre-processing and Noise Rejection

Considering the ECG sampling frequency is essential. The insufficient sampling frequency may negatively affect the resolution and the signal-to-noise ratio of the R-R time series [41, 42]. The UCD and the Apnea-ECG databases have less sampling frequency than the specified 250Hz value of the American National Standard Institute (ANSI), yet they are good benchmarks for the evaluation of automatic OSA detection methods. We have assumed that their subjects are carefully selected so that exceptions, where their sampling frequencies are insufficient for representing ECG behavior, are deleted [42]. The ECG signals of the exclusive database are also down-sampled to 250Hz.

To avoid the aliasing effects of non-integer fractional down-sampling, equating the UCD and the Apnea-ECG sampling frequencies is avoided [43]. For de-trending and noise rejection, the decimated lifting wavelet transform (DWT) algorithm [44] is employed [13]. The Daubechies (D4) wavelet is used with seven levels of decomposition. The R-R time series is extracted by the famous and robust method of Hamilton-Tompkins [45, 46]. Impulses more or less than 20% distant to the last normal R-R interval, those with more than 30% values in the R-S difference or with the negative R-S difference values are assumed to be a sign of ectopic or abnormal beat and omitted; the resulting signal is called the R-R tachogram [38].

Table 1. The SpO₂ features in each 1-minute frame: $\{spo2_i\}_{i=1}^{60}$

Name/ Definition
The minimum value of the frame
The average value of the frame
The standard deviation of the frame
Sequential correlation coefficients [20]
Sequential mutual information [52]
Average value crossing points
The absolute value of the slope of the line fitted over SpO ₂ [20]
y-Intercept value of the line fitted over SpO ₂ [20]
Approximate entropy [53]
Sample entropy [53]
Lempel-Zive complexity measure [54]
Central tendency measure (CTM _r) (r=0.25, 0.75, 0.5, 1)[54]
Delta measure (Δ) [30]
Baseline [22]
odi2, odi3, odi4: The number of 2%,3%, and 4% desaturations to the baseline [30]
ODI _{xy} : The number of desaturations more than or equal to x% lasting for y seconds [30]
ODIS _x : The number of desaturations more than or equal to x% [22]
Time elapsed under saturation level x (%tsax); x=70, 80,85, 90, 95) [30]

2.2. Feature Extraction

We consider values below 50% and fluctuations more than 40% in two consecutive samples of SpO2 signal (in the sampling period of 1s) artifacts [16, 19]. We eliminate these values and their corresponding values of other PSG signals from the records (2 minutes of the Apnea-ECG database, 37 minutes of the UCD database, and 78 minutes of the exclusive database, totally equal to 1.9% of available data). The resulting signal is divided into non-overlapping 1-minute frames and is used for feature extraction. Table 1 summarizes the SpO2 features.

We process the ECG signal in 1-minute time windows. The R-R tachogram is extracted from ECG. It is not a result of uniform ECG sampling. The points of this time series are scattered non-uniformly across the time axis based on the time interval of consecutive beats. In frequency analysis of ECG signal, this crucial fact is usually ignored. The pre-assumption of the fast Fourier transform (FFT) is the uniform sampling of the signal under analysis; hence the FFT-based frequency analysis of the R-R tachogram and its dependents like the ECG-derived respiration (EDR) are not appropriate. Frequency analysis tools needless of the uniform sampling assumption like the Lomb-Scargle periodogram are good candidates for calculating quantities related to the heart rate variability (HRV) [50].

The EDR is extracted by the T wave duration method [51, 52] in the UCD and ECG-Apnea databases. We calculate the EDR with the help of the area under the QRS graph [53] in our exclusive database.

We use the Lomb-Scargle periodogram and the DWT with Daubechies (D4) wavelet (with 18 levels of decomposition) to extract frequency-domain features of the R-R tachogram, and the EDR signals [44]. The ECG features are categorized as the time-domain, and the frequency-domain features in tables 2, 3 and 4.

Table 2. The time-domain ECG features

The R-R tachogram: $R(rr_{t_m}) = \{rr_i\}_{i=rr_{t_1}}^{rr_{t_m}}$, the EDR: $EDR(q) = \{edr_i\}_{i=1}^q$

Definition	Name
$\bar{rr}_t = \frac{1}{m} \sum_{i=1}^m rr_{t_i}$	Time window mid-time
M	length ECG
$\bar{rr} = \frac{1}{m} \sum_{i=1}^m rr_i$	Average beat [115]
$NN50v1 = \sum_{i=2}^m U(rr_i - rr_{i+1} - 50ms)$ U(.): step function	NN50-version 1 [115]
$NN50v2 = \sum_{i=1}^{m-1} U(rr_{i+1} - rr_i - 50ms)$ U(.): step function	NN50-version 2 [115]
$pNN50v1 = \frac{NN50v1}{m}$	pNN50-version 1 [115]
$pNN50v2 = \frac{NN50v2}{m}$	pNN50-version 2 [115]
$S_{rr} = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (rr_i - \bar{rr})^2}$	Tachogram standard deviation
$S_{DSD} = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (rd_i - \bar{rd})^2}$ $rd_i = rr_{i+1} - rr_i$, $\bar{rd} = \frac{1}{m-1} \sum_{i=1}^{m-1} rd_i$	S_{DSD} [115]

$RMSSD = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m-1} rd_i^2}$	RMSSD [115]
$r_k = \frac{\sum_{i=1}^m (rr_i - \bar{rr})(rr_{i+k} - \bar{rr})}{\sum_{i=1}^m (rr_i - \bar{rr})^2}$	Sequential correlation coefficients [115]
$MI_k = \hat{I}(\{rr_i\}; \{rr_{i+k}\})$ $= \sum_{i=1}^m P_n(\{rr_i\}, \{rr_{i+k}\}) \log \frac{P_n(\{rr_i\}, \{rr_{i+k}\})}{P_n(\{rr_i\})P_n(\{rr_{i+k}\})}$ P_n : Probability distribution function	Sequential mutual information [319]
$AT_k = \frac{E((N_{i+1}(k) - N_i(k))^2)}{2E(N_{i+1}(k))}$ $N_i(k)$: Number of beats in the i^{th} section of a k -second signal	Allan Factor [124]
$NEP_k = \frac{1}{m-2} \sum_{i=2}^{m-1} (1 - U((rr_i - rr_{i-1})(rr_{i+1} - rr_i)))$	Number of Extreme Points [116]
$\bar{edr} = \frac{1}{q} \sum_{i=1}^q edr_i$	Average EDR
$S_{edr} = \sqrt{\frac{1}{q-1} \sum_{i=1}^q (edr_i - \bar{edr})^2}$	Standard Deviation EDR

Table 3. The frequency-domain features of the R-R tachogram: $R(rr_{t_m}) = \{rr_i\}_{i=rr_{t_1}}^{rr_{t_m}}$

Definition	Name
$S^2_{D_{rr}^k} = \sum_{i=1}^{I_{rr,k}} (d_{rr,i}^k - \bar{d}_{rr}^k)^2$ $\bar{d}_{rr}^k = \frac{1}{I_{rr,k}} \sum_{i=1}^{I_{rr,k}} d_{rr,i}^k$	Sample deviation of $\{D_{rr}^k\}_{k=2}^{17}$
$S^2_{D_{rr}^{LF}} = \sum_{i=1}^{I_{rr,LF}} (d_{rr,i}^{LF} - \bar{d}_{rr}^{LF})^2$	Sample deviation of $\{D_{rr}^k\}_{k=2}^{17}$ (LF band)
$S^2_{D_{rr}^{HF}} = \sum_{i=1}^{I_{rr,VLF}} (d_{rr,i}^{HF} - \bar{d}_{rr}^{HF})^2$	Sample deviation of $\{D_{rr}^k\}_{k=2}^{17}$ (HF band)
$P_{rr}^{VLF} = \int_{2\pi \times 0.04}^{2\pi \times 0.15} P_{rr}(\omega) d\omega$ $P_{rr}(\omega)$: Lomb-Scargel periodogram [348,86]	HRV Power spectrum (LF band)
$P_{rr}^{HF} = \int_{2\pi \times 0.15}^{2\pi \times 0.4} P_{rr}(\omega) d\omega$	HRV Power spectrum (HF band)
$LF/HF = P_{rr}^{LF} / P_{rr}^{HF}$	LF-HF power ratio in the HRV spectrum
$P_{rr}(\omega) _{2\pi \times 0.04}^{2\pi \times 0.4}$	Lomb-Scargel periodogram samples in LF-HF band
$\omega_{resp} = \operatorname{argmax}(P_{rr}(\omega) _{2\pi \times 0.15}^{2\pi \times 0.4})$	Estimated respiration frequency (Dominant HF-band frequency of HRV) [86]
$respMag = \max(P_{rr}(\omega) _{2\pi \times 0.15}^{2\pi \times 0.4}) = P_{rr}(\omega_{resp})$	Power at the dominant HF-band frequency of

	HRV
$respProb = Prob(P_{rr}(\omega_{resp}))$	Probability of estimated respiration frequency occurrence with power $P_{rr}(\omega_{resp})$
$\omega_{ProbMax} = argmax(Prob(P_{rr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.04}}))$	Most probable frequency of the HRV spectrum
$ProbMax = max(Prob(P_{rr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.04}}))$ $= Prob(P_{rr}(\omega_{ProbMax}))$	Probability of $\omega_{ProbMax}$ occurrence with power $P_{rr}(\omega_{ProbMax})$
$ProbMaxMag = P_{rr}(\omega_{ProbMax})$	Power of the HRV spectrum at $\omega_{ProbMax}$

Table 4. The frequency-domain features of the EDR: $EDR(q) = \{edr_i\}_{i=1}^q$

Definition	Name
$S^2_{D_{edr}^k} = \sum_{i=1}^{I_{edr,k}} (d_{edr,i}^k - \overline{d_{edr}^k})^2$ $\overline{d_{edr}^k} = \frac{1}{I_{edr,k}} \sum_{i=1}^{I_{edr,k}} d_{edr,i}^k$	Sample deviation of $\{D_{edr}^k\}_{k=2}^{17}$
$S^2_{D_{edr}^{LF}} = \sum_{i=1}^{I_{edr,LF}} (d_{edr,i}^{LF} - \overline{d_{edr}^{LF}})^2$	Sample deviation of $\{D_{edr}^k\}_{k=5}^{17}$ (LF band)
$S^2_{D_{edr}^{HF}} = \sum_{i=1}^{I_{edr,VLF}} (d_{edr,i}^{HF} - \overline{d_{edr}^{HF}})^2$	Sample deviation of $\{D_{edr}^k\}_{k=2}^4$ (HF band)
$P_{edr}^{VLF} = \int_{\frac{2\pi \times 0.04}{2\pi \times 0.15}}^{2\pi \times 0.15} P_{edr}(\omega) d\omega$	EDR Power spectrum (LF band)
$P_{edr}^{HF} = \int_{\frac{2\pi \times 0.15}{2\pi \times 0.04}}^{2\pi \times 0.04} P_{edr}(\omega) d\omega$	EDR Power spectrum (HF band)
$LF/HF_{edr} = P_{edr}^{LF} / P_{edr}^{HF}$	LF-HF power ratio in the EDR spectrum
$P_{edr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.04}}$	Lomb-Scargel periodogram samples in LF-HF band
$\omega_{edr-resp} = argmax(P_{edr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.15}})$	Dominant HF-band frequency of the EDR
$respMag_{edr} = max(P_{edr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.15}})$ $= P_{edr}(\omega_{edr-resp})$	Power at the dominant HF-band frequency of the EDR
$respProb_{edr} = Prob(P_{edr}(\omega_{edr-resp}))$	Probability of $\omega_{edr-resp}$ occurrence with power $P_{edr}(\omega_{edr-resp})$
$\omega_{edr-ProbMax} = argmax(Prob(P_{edr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.04}}))$	Most probable frequency of the EDR spectrum
$ProbMax_{edr} = max(Prob(P_{edr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.04}}))$ $= Prob(P_{edr}(\omega_{edr-ProbMax}))$	Probability of $\omega_{edr-ProbMax}$ occurrence with power $P_{edr}(\omega_{edr-ProbMax})$
$ProbMaxMag_{edr} = P_{edr}(\omega_{edr-ProbMax})$	Power of the HRV spectrum at $\omega_{edr-ProbMax}$

2.3. Feature Reduction

Most automatic OSA detection methods [11-13, 16-19, 27, 29] use no feature reduction or employ linear dependency and correlation-based strategies or principal component analysis (PCA) for feature selection. Dependency and mutual information (MI) proved to outperform linear methods of feature selection, especially in respiratory event detection [14, 31]. Feature selection can be performed by individual analysis of each feature [13, 21, 26]. It is also possible to define a measure to evaluate a subset of features [14, 16]. The first method speculates the inter-relations among features but, the second method searches for features with both the tightest relations with the class label and the loosest interaction with each other. We use the second strategy for feature reduction.

To calculate the mutual interactions, we consider MI rather than a simple statistical correlation. We select the features which have the highest MI with the class label (normal or apnoeic) and the least MI with each other. The approach to search the feature space is forward feature selection. In this approach, the subset of selected features is gradually built by adding single features to an initial null set [14, 54].

2.4. Classification

We employ nine classifiers in this study; support vector machines (SVM) [55], K nearest neighbors (KNN) [60], decision table [56], C4.5 [57] decision tree, reduced-error pruning tree (REPT) [58], functional trees [59], the meta-algorithm of adaptive boosting accompanied with the simple classifier of decision stump [60], and the meta-algorithm of bagging along with the alternating decision tree (ADT) [61]. The meta-algorithms make a new data set out of the primary data set and devise a new classifier for each set in one trial. These trials are repeated T times, and eventually, the results of the T classifiers are combined to achieve a more accurate result.

In this study, four classifier combination methods are also performed on a group of three binary classifiers. Combination methods are max probability, average probability, the product of probability, and majority voting [16].

3. RESULTS

Table 5 demonstrates the selected features employing the MI measure. According to table 5, as the number of database subjects increases, the number of selected features also increases. There are several similarities between the selected measures; fewer ECG features are among the selected ones, mostly the time domain ECG features. This result is consistent with the previously published reports. Most of the selected features are based on the SpO2 signal, which indicates their power for the OSA detection. However, simultaneous use of the ECG and the SPO2 features enhances the performance of the OSA detection method [16].

Table 5. The selected features through forward feature selection based on the MI criterion. Name and definition of features stated in tables 1 to 4

Number	Selected features	Database
20	MI_3 , $spo2_{min}$, NEP_1 , S_{spo2} , $MI_{spo2,1}$, Δ , LZ_{down} , $odi4$, $CTM_{0.5}$, $ODI55$, $tsa80$, $tsa85$, $tsa90$, $S^2_{D_{rr}^4}$, P_{rr}^{HF} , $S^2_{D_{edr}^6}$, P_{edr}^{LF} , 'samples 13 th and 55 th of $P_{rr}(\omega)$ sample 4 th of $P_{edr}(\omega)$	UCD
18	S_{spo2} , $MI_{spo2,1}$, Δ , LZC_{up} , $CTM_{0.25}$, $CTM_{0.5}$, $ODI55$, $tsa80$, $tsa85$, $tsa90$, $S^2_{D_{rr}^4}$, $S^2_{D_{edr}^6}$, P_{edr}^{LF} , samples 11 th , 18 th , 22 th and 55 th of $P_{rr}(\omega)$ and sample 4 th of $P_{edr}(\omega)$	Apnea-ECG
29	$Spo2_{min}$, $\overline{spo2}$, S_{spo2} , $r_{spo2,2}$, ZC , $ApEn$, $SpEn$, LZC_{up} , $MI_{spo2,3}$, $MI_{spo2,4}$, Δ , $ODIS4$, $ODI23$, $ODI25$, $ODI31$, $ODI35$, $ODI51$, $ODI53$, $ODI55$, $odi3$, $odi4$, $odi5$, $tsa95$, $tsa85$, $tsa80$, $CTM_{0.5}$, $CTM_{0.75}$, CTM_1 , 'sample 4 th of $P_{edr}(\omega)$	Exclusive database

Table 6 illustrates the performance of our real-time detection method in each of the databases. We obtain the results from a system equipped with Windows 10 Pro, version 1511, the Intel processor Core i7CPU M640@2.8GHz and a RAM of 4GB. All the classifiers are realized in Java language. Evaluation is 10-fold cross-validation.

In some references, only the classifier's training time is reported [16]. This parameter is not enough to represent the total computational burden of the suggested method. In some previous works, the processing time is reported for a specified number of samples [14]. In our study, "the processing time for a fixed number of data samples" is not an accurate measure since several databases with different ECG sampling rates are observed.

Table 6. The performance of the suggested detection method in each of the databases: DT (Decision Table), REPT (Reduced-Error Pruning Tree), FT (Functional tree), AB+DS (Adaptive boosting + decision stump), B+ REPT (Bagging + REPT), B+ADT (bagging + alternating decision tree), AECG (Apnea-ECG database), EX (Exclusive database). Maximums in each column are shaded.

Processing time for 10 frames			Accuracy (%)			Specificity (%)			Sensitivity (%)			Classifier
EX	AECG	UCD	EX	AECG	UCD	EX	AECG	UCD	EX	AECG	UCD	
19	11.8	11.9	88.3	95.3	82	91	89.8	93	80.9	96.68	81.02	SVM
7	2.98	2.09	82.9	90.4	82	84.7	94	83	80.01	89	80.5	KNN
5.68	3.001	2.503	82	83.7	82	83	84.9	82	82.9	83	82.9	DT
4.001	1.45	1.076	82	85.6	81.7	86.1	89	85	73	82.1	72	C4.5
2.32	1.045	1.002	84.6	91.6	83.6	84.9	92.6	84	82.9	83.5	81.5	REPT
9.867	4.7	4.345	80	88.8	79.8	82	90.7	81.7	73	81.4	71.5	FT
2.383	1.32	1.205	92.6	87.3	79	93.3	79.3	78	89.9	92.6	88	AB+DS
4.794	2.97	2.164	88.5	91	85	89.9	92.2	86.3	82.1	89	81.03	B+REPT
29.9	15	13.98	85.6	95	84.5	85	95	83	86.78	89.9	85	B+ADT
18	9	8.99	55.1	63	57	55.8	57	54.3	55.01	65	59	SOM
10	5.7	4.897	35.6	38.6	34.5	33	37	33	38.4	40.1	37.3	K-means

Observing the processing time in table 6 reveals that the parameter value does not exceed 1s in the UCD and Apnea-ECG databases and 2s in our exclusive database. These margins are the minimum time needed for pre-processing and feature extraction at the specified sampling frequencies. Smaller values for processing times belong to the Apnea-ECG database with the lowest number of data points. The processing time for our exclusive database is the highest of all, nearly two times the minimum value. Regarding this quantity, two classifiers have the highest computational burden; the ADT and the SVM. The processing time for the SVM is more than two times higher than the others'. For the real-time OSA detection, these computationally intensive classifiers are not chosen despite their high classification ability.

Accuracy, sensitivity, and specificity in all the databases are satisfactory but, slightly better in the Apnea-ECG database compared to the others. The two unsupervised classifiers (the SOM and the K-means) do not exhibit acceptable results. Best sensitivity, but the worst specificity/accuracy belongs to adaptive boosting accompanied with the decision stump. On the other hand, bagging along with REPT achieves the best accuracy and specificity at the price of degrading sensitivity.

To reach a method with acceptable sensitivity and specificity, the combination routines declared in section 2.4 are used to fuse a group of three classifiers. Because the “boosting with the decision stump” and the “bagging along with the REPT” have better performances than others, they are the two fixed members of the group. The third member is chosen from the rest of the classifiers. We exclude the SVM and the “bagging with ADT” due to excessive computational load, so five options remain. These classifiers shape five different classifier groups to be fused. The classifier combination results are reported in tables 7 to 9.

According to tables 7 to 9, performance is nearly equal in all databases (slightly better performance for the Apnea-ECG database). Combining the classifiers, balances the performance measures in values around 80%. The most successful combination happened when the third group member is the KNN or the decision tree. In these cases, all the measures of performance, including sensitivity, specificity, and accuracy, have achieved values of more than 85%. These results outperform all the suggested methods to date [13, 14, 16, 19, 32, 35]. The principal difference between the KNN and the decision tree lies in their nature. KNN benefits from slow, moment-based training. It is appropriate for subject-dependant applications, in which models are built and tested with the same data. In subject-dependant applications each classifier model should be trained (i.e. updated) with the user data before utilization. On the other hand, the decision table is suitable for subject-independent applications where the classifier model is trained with a database of several subjects before being tested by the user.

Surveying the processing time shows that this quantity is approximately equal to the sum of the processing time needed for each classifier of the group. There is no distinguished difference between different combination routines. It is worth saying that combination methods based on probability need the sensitivity and the specificity of the classifier to weigh their decisions. This issue entails a more complex online realization than that of majority voting. Therefore, in online realization, the majority voting method will suffice.

Table 7. The performance of the suggested classifier combination detection method in the UCD database. Three classifiers are combined with four different methods (MP: Maximum probability, PP: Probability product, AP: Average probability, MV: Majority voting). Other abbreviations are similar to table 6. The two highest values in each column are shaded.

Processing time for 10 frames				Accuracy (%)				Specificity (%)				Sensitivity (%)				3 rd classifier
MV	AP	PP	M P	M V	AP	PP	M P	M V	AP	PP	M P	M V	AP	P P	M P	
4.36	4.68	4.47	4.4	85.28	86.12	86.2	86.12	85.25	86.03	86.16	86.07	87.55	87.41	87.19	85.87	KNN
4.55	4.65	4.76	4.869	85	85.68	85.7	85.64	84.16	85.35	85.42	85.47	87.61	86.68	86.57	86.14	DT
4	3.92	3.79	3.963	81.81	82.12	82.17	82.02	81.25	82.03	82.16	82.07	83.55	82.41	82.19	81.87	C4.5
3.65	3.56	3.39	3.245	81	81.68	81.70	81.64	80.16	81.35	81.42	81.47	83.61	82.68	82.57	82.14	REP T
5.21	5.34	5.63	5.56	81.03	80.95	80.98	80.96	80.43	80.48	80.57	80.69	82.9	82.41	82.25	81.82	FT

Table 8. The performance of the suggested classifier combination detection method in the Apnea-ECG database. Three classifiers are combined with four different methods (MP: Maximum probability, PP: Probability product, AP: Average probability, MV: Majority voting). Other abbreviations are similar to table 6. The two highest values in each column are shaded.

Processing time for 10 frames				Accuracy (%)				Specificity (%)				Sensitivity (%)				3 rd classifier
MV	AP	PP	MP	MV	AP	PP	MP	MV	AP	PP	MP	MV	AP	PP	MP	
5.68	5.634	5.555	5.29	85.38	86.2	6.227	86.15	85.34	86.61	86.23	86.17	87.6	87.5	87.2	86	KN N
5.23	5.125	5.34	5.291	85.02	85.7	8.58	85.2	84.2	85.39	85.48	85.5	86.70	86.73	86.6	86.23	DT
3.99	3.7	3.65	3.49	82.4	82.25	8.22	82.23	82.33	82.21	82.2	82.1	83.6	82.5	82.2	81.94	C4.5
3.025	3.068	3.128	3.11	81	81.71	1.75	81.7	80.2	81.38	81.49	81.5	83.69	82.7	82.6	82.15	REP T
5.969	5.79	5.87	5.9	81.1	81.2	1.01	81	80.57	80.6	80.7	81	83	82.5	82.31	81.91	FT

Table 9. The performance of the suggested classifier combination detection method in the exclusive database. Three classifiers are combined with four different methods (MP: Maximum probability, PP: Probability product, AP: Average probability, MV: Majority voting). Other abbreviations are similar to table 6. The two highest values in each column are shaded.

Processing time for 10 frames				Accuracy (%)				Specificity (%)				Sensitivity (%)				3 rd classifier
MV	AP	PP	MP	MV	AP	PP	MP	MV	AP	PP	MP	MV	AP	PP	MP	
12.05	11.81	11.6	12	85.32	86.03	86.13	86	85.24	86.6	86.11	86	87.23	87.35	87.1	85.67	KN N
10.43	10.54	10.68	10.56	84.9	85.48	85.7	85.6	84.14	85.30	85.34	85.5	87.5	86.65	86.6	86	DT
9.34	9.47	9.24	9.04	81.78	82.10	82.13	81.95	81.20	82.8	82.14	82.01	83.51	82.13	82.17	81.85	C4.5
7.349	7.367	7.489	7.32	82.1	81.68	81.67	81.64	81.55	81.33	81.43	81.46	83.6	82.7	82.55	82.1	REP T
15.004	14.96	14.62	14.86	81.01	80.87	80.93	80.92	80.32	80.6	80.44	80.59	82.8	82.3	82.15	81.72	FT

4. CONCLUSIONS

In this study, several configurations for online detection of the OSA are suggested. The advantages of the proposed method are: exploiting only two channels of biological signals, automatic and real-time detection, and uniform acceptable performance over several databases (over 85%). To date, no other study has achieved all these merits together. Acceptable performance in well-known databases is due to classifiers that do not possess database-related parameters (e.g. sampling frequency of signals). The classifiers have covered deficiencies of each other in a combinational configuration. To reach the best result, the most successful classifiers are combined in groups of three members with four different combination methods. The features are also calculated and selected considering generality; in frequency-domain analysis, the refined Lomb-Scargle periodogram is used to care for the inherent non-uniform sampling of the R-R tachograms and unequal sampling frequency of the ECG signal in different databases [50]. Feature selection is based on the MI. The MI measure considers non-linear correlations among features and selects effective features to decrease the computational burden of the classifiers and avoid over-fitting problems.

On the other hand, the MI feature reduction has an important impact on the family of decision tree classifiers. MI-based feature selection accompanied by decision tree classifiers, avoids the classifier sensitivity to MI-biased estimates. In other words, the decision-tree classifiers may be misled by a fake replica of a feature with more marginal samples and higher maximum entropy value [62]. Selection of the more appropriate feature with an entropy-normalised MI estimator is helpful [62, 63].

ACKNOWLEDGEMENTS

The author appreciates the cooperation of the Ibn-e-Sina Hospital sleep laboratory at Mashhad University of Medical Sciences.

REFERENCES

- [1] C. V. Senaratna, J. L. Perret et al, "Prevalence of obstructive sleep apnea in the general population: A systematic review", *Sleep Medicine Reviews*, vol. 34, pp. 70-81, 2017. ISSN 1087-0792, Available: <https://doi.org/10.1016/j.smrv.2016.07.002>.
- [2] O. Berdina, I. Madaeva, and L. Rychkova, "Obstructive Sleep Apnea and COVID-19 Infection Comorbidity: Analysis of the Problem in the Age Aspect", *International Journal of Biomedicine*, vol. 10, no. 4, pp. 312-315, 2020. Available: 10.21103/article10(4)_ra1.
- [3] J. Remmers, W. deGroot, E. Sauerland and A. Anch, "Pathogenesis of upper airway occlusion during sleep", *Journal of Applied Physiology*, vol. 44, no. 6, pp. 931-938, 1978. Available: 10.1152/jappl.1978.44.6.931.
- [4] B. Dredla and P. Castillo, "Cardiovascular Consequences of Obstructive Sleep Apnea", *Current Cardiology Reports*, vol. 21, no. 11, 2019. Available: 10.1007/s11886-019-1228-3.
- [5] B. Phillips, "Association of Sleep Apnea and Type II Diabetes: A Population-Based Study", *Yearbook of Pulmonary Disease*, vol. 2007, pp. 249-251, 2007. Available: 10.1016/s8756-3452(08)70452-3.
- [6] W. Akberzie, S. Hesselbacher, I. Aiyer, S. Surani and Z. Surani, "The Prevalence of Anxiety and Depression Symptoms in Obstructive Sleep Apnea", *Cureus*, 2020. Available: 10.7759/cureus.11203.
- [7] W. McNicholas, "Diagnosis of Obstructive Sleep Apnea in Adults", *Proceedings of the American Thoracic Society*, vol. 5, no. 2, pp. 154-160, 2008.
- [8] *International classification of sleep disorders*. Darien, Ill.: American Acad. of Sleep Medicine, 4th ed. New York: Westchester, III; 2014.
- [9] T. Lee-Chiong, *Sleep*. Hoboken, N.J.: Wiley-Liss, 2006.

- [10] V. Kapur et al., "The Medical Cost of Undiagnosed Sleep Apnea", *Sleep*, vol. 22, no. 6, pp. 749-755, 1999. Available: 10.1093/sleep/22.6.749.
- [11] U. R. Abeyratne, V. Swarnkar, C. Hukins, and B. Duce, "Interhemispheric Asynchrony Correlates With Severity of Respiratory Disturbance Index in Patients With Sleep Apnea," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2947-2955, 2010.
- [12] Kim PY, McCarty DE, Wang L, Frilot C, Chesson AL, Marino AA. Two-group classification of patients with obstructive sleep apnea based on analysis of brain recurrence. *Clinical Neurophysiology*. 2014 Jun 30;125(6):1174-81.
- [13] Bsoul M, Minn H, Tamil L. Apnea MedAssist: real-time sleep apnea monitor using single-lead ECG. *IEEE Transactions on Information Technology in Biomedicine*. 2011 May;15(3):416-27.
- [14] Nguyen HD, Wilkins BA, Cheng Q, Benjamin BA. An online sleep apnea detection method based on recurrence quantification analysis. *IEEE Journal of Biomedical and Health Informatics*. 2014 Jul;18(4):1285-93.
- [15] H. M. Al-Angari, and Alan V. Sahakian, "Automated Recognition of Obstructive Sleep Apnea Syndrome Using Support Vector Machine Classifier," *IEEE Trans. Inf. Tech. Biomed.*, vol. 16, no. 3, pp. 463-468, 2012.
- [16] Xie B, Minn H. Real-time sleep apnea detection by classifier combination. *IEEE Transactions on Information Technology in Biomedicine*. 2012 May;16(3):469-77.
- [17] Sannino G, De Falco I, De Pietro G. Monitoring obstructive sleep apnea by means of a real-time mobile system based on the automatic extraction of sets of rules through differential evolution. *Journal of biomedical informatics*. 2014 Jun 30; 49:84-100.
- [18] J. V. Marcos et al, "Automated Prediction of the Apnea-Hypopnea Index from Nocturnal Oximetry Recordings," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 1, pp. 141-149, 2012.
- [19] Koley BL, Dey D. On-Line Detection of Apnea/Hypopnea Events Using SpO2 Signal: A Rule-Based Approach Employing Binary Classifier Models. *IEEE Journal of Biomedical and Health Informatics*. 2014 Jan;18(1):231-9.
- [20] Koley BL, Dey D. Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary SVM classifiers. *Measurement*. 2013 Aug 31;46(7):2082-92.
- [21] Koley BL, Dey D. Real-time adaptive apnea and hypopnea event detection methodology for portable sleep apnea monitoring devices. *IEEE Transactions on Biomedical Engineering*. 2013 Dec;60(12):3354-63.
- [22] S. I. Rathnayake, I. A. Wood, U. R. Abeyratne, and C. Hukins, "Nonlinear features for single-channel diagnosis of sleep-disordered breathing diseases," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1973-1981, Aug. 2010.
- [23] Sanchez-Morillo D, Lopez-Gordo MA, Leon A. Novel multiclass classification for home-based diagnosis of sleep apnea/hypopnea syndrome. *Expert Systems with Applications*. 2014 Mar 31;41(4):1654-62.
- [24] Azarbarzin, Ali, and Zahra MK Moussavi. "Automatic and unsupervised snore sound extraction from respiratory sound signals." *IEEE Transactions on Biomedical Engineering* 58.5 (2011): 1156-1162.
- [25] E. Goldshtein, A. Tarasiuk, and Y. Zigel, "Automatic Detection of Obstructive Sleep Apnea Using Speech Signals," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1373-1382, Feb. 2011.
- [26] Burgos A, Goni A, Illarramendi A, Bermudez J. Real-time detection of apneas on a PDA. *IEEE Transactions on Information Technology in Biomedicine*. 2010 Jul;14(4):995-1002.
- [27] Behar J, Roebuck A, Shahid M, Daly J, Hallack A, Palmius N, Stradling J, Clifford GD. SleepAp: An automated obstructive sleep apnoea screening application for smartphones. *IEEE journal of biomedical and health informatics*. 2015 Jan;19(1):325-31.
- [28] ZhaoD, WangY, WangQ, WangX. Comparative analysis of different characteristics of automatic sleep stages. *Comput Methods Programs Biomed* 2019; 175:53-72
- [29] Carrubba S, Kim PY, McCarty DE, Chesson Jr AL, Frilot C, Marino AA. Continuous EEG-based dynamic markers for sleep depth and phasic events. *J. Neurosci. Meth.* 2012; 208:1-9.
- [30] Ireneusz Jabłoński, "Modern Methods for the Description of Complex Couplings in the Neurophysiology of Respiration," *IEEE Sensors Journal*, vol. 13, no. 9, pp. 3182-3192, Sep 2013.
- [31] H. Ghandeharioun, F. Rezaeitalab, and R. Lotfi, "Analysis of respiratory events in obstructive sleep apnea syndrome: Inter-relations and association to simple nocturnal features", *Revista Portuguesa de Pneumologia (English Edition)*, vol. 22, no. 2, pp. 86-92, 2016. Available: 10.1016/j.rppnen.2015.09.008.

- [32] ErdenebayarU, KimYJ, ParkJ-U, JooEY, LeeK-J. Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram. *Comput. Methods Programs Biomed.* 2019; 180:105001.
- [33] DaldalN, CömertZ, PolatK. Automatic determination of digital modulation types with different noises using convolutional neural network based on time–frequency information. *Appl. Soft Comput.* J 2019.
- [34] ChoiSH, YoonH, KimHS, KimHB, KwonHB, OhSM, et al. Real-time apnea-hypopnea event detection during sleep by convolutional neural networks. *Comput. Biol. Med.* 2018; 100:123–31.
- [35] F. Bozkurt, M.K. Uçar, M.R. Bozkurt, C. Bilgin, Detection of Abnormal Respiratory Events with Single Channel ECG and Hybrid Machine Learning Model in Patients with Obstructive Sleep Apnea, *IRBM*, Volume 41, Issue 5, 2020, Pages 241-251.
- [36] St. Vincent’s University Hospital/University College Dublin Sleep Apnea Database. (2008). Available: <http://www.physionet.org/pn3/ucddb/>
- [37] “CinC Challenge 2000 data sets: Data for development and evaluation of ECG-based apnea detectors,” (2000). Available: <http://www.physionet.org/physiobank/database/apnea-ecg/>
- [38] A. Roebuck, V. Monasterio, E. Geder, M. Osipov, J. Behar, A. Malhotra, T. Penzel and G. Clifford, "A review of signals used in sleep analysis", *Physiological Measurement*, vol. 35, no. 1, pp. R1-R57, 2013.
- [39] Kuna S T, Benca R, Kushida C A, Walsh J, Younes M, Staley B, Hanlon A, Pack A I, Pien G W and Malhotra A 2013 Agreement in computer-assisted manual scoring of polysomnograms across sleep centres, *Sleep* 36 583–9
- [40] Behar J, Roebuck A, Shahid M, Daly J, Hallack A, Palmius N, Stradling JR, Clifford GD. SleepAp: An automated obstructive sleep apnoea screening application for smartphones, *Computing in Cardiology* 2013 Sep 22 (pp. 257-260).
- [41] Malik M., Camm A.J. (eds.): *Heart Rate Variability*, Armonk, N.Y. Futura Pub. Co. Inc., 1995.
- [42] Abboud S., Barnea O.: Errors due to sampling frequency of electrocardiogram in spectral analysis of heart rate signals with low variability, *Computers in Cardiology*, pp 461-463, 1995.
- [43] Oppenheim, Alan V., Willsky, Alan S., Navab, S. Hamid. "Digital signal processing." Printice-Hall, London (2000). ISBN:0-13-651175-9.
- [44] Daubechies, Ingrid. "Ten lectures on wavelets, vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics." (1992).
- [45] Hamilton P., Tompkins W.: Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database, *IEEE Transactions on Biomedical Engineering*, vol. BME-33, NO. 12. 1986.
- [46] Pan J., Tompkins W.: A real-time QRS detection algorithm, *IEEE Transactions on Biomedical Engineering*, vol. BME-32 NO. 3. 1985.
- [47] Darbellay, Georges A., and Igor Vajda. "Estimation of the information by an adaptive partitioning of the observation space." *IEEE Transactions on Information Theory* 45.4 (1999): 1315-1321.
- [48] R. Hornero, D. Alvarez, D. Abasolo, F. del Campo, and C. Zamarron, "Utility of approximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, pp. 107-113, Jan. 2007.
- [49] D. Alvarez, R. Hornero, D. Ab'asolo, F. Campo, and C. Zamarr'on, "Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection," *Physiol. Meas.*, vol. 27, pp. 399–412, 2006.
- [50] Clifford GD. Signal processing methods for heart rate variability (Doctoral dissertation, Department of Engineering Science, University of Oxford, September 2002).
- [51] G. D. Furman, Z. Shinar, A. Baharav, and S. Akselrod, "Electrocardiogram derived respiration during sleep," *Comput. Cardiol.*, vol. 32, pp. 351–354, 2005.
- [52] Raymond, B., Cayton, R. M., Bates, R. A., & Chappell, M. J., "Screening for Obstructive Sleep Apnoea Based on the Electrocardiogram—The Computers in Cardiology Challenge," *Proc. Computers in Cardiology*, Vol. 27, IEEE Press, 2000, pp. 267–270.
- [53] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multi-lead ECGs," *Comput. Cardiol.*, vol. 12, pp. 113–116, 1985.
- [54] P. A. Estevez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Feb. 2009.
- [55] Murphy KP. *Machine learning: a probabilistic perspective*. MIT Press; 2012 Sep 7.
- [56] N. Landwehr, M. Hall, and E. Frank. (2005, May). Logistic model trees. *Mach. Learn.* 59, pp. 161–205.

- [57] R. Quinlan, C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann, 1993.
- [58] Mingers, John. "An empirical comparison of pruning methods for decision tree induction." Machine learning 4.2 (1989): 227-243.
- [59] Gama, Joao. "Functional trees." Machine Learning, vol. 55, no. 3, pp. 219-250, 2004.
- [60] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123–140, 1996.
- [61] Y. Freund and L. Mason, "The alternating decision tree algorithm," in Proc. 16th Int. Conf. Mach. Learning, 1999, pp. 124–133.
- [62] J. R. Quinlan, "Induction of decision trees," Mach. Learn., vol. 1, pp. 81–106, 1986.
- [63] Paninski, Liam. "Estimation of entropy and mutual information." Neural computation 15.6 (2003): 1191-1253.

AUTHOR

Dr. Hosna Ghandeharioun received her B.S. degree in electrical engineering from Ferdowsi University of Mashhad (FUM), Iran, in 2003 and an M.S. degree in Biomedical Engineering from Iran University of Science & Technology, Iran in 2006. She received her Ph. D. degree in electrical engineering from FUM in 2016. Now She works as an assistant professor at the electrical and biomedical Engineering dept. of Khorasan Institute of Higher Educations. Her research interests are biological signal processing and mobile health.

