Feature Selection for Obstructive Sleep Apnea Recognition

Araştırma Makalesi/Research Article

Mustafa SERT

Department of Computer Engineering, Başkent University, Ankara, Türkiye
msert@baskent.edu.tr
(Geliş/Received:03.09.2019; Kabul/Accepted:16.10.2019)
DOI: 10.17671/gazibtd.615014

***Abstract*—** Obstructive sleep apnea (OSA) is a kind of sleep disorder and it is described by breathing irregularity during sleep. This disorder may lead to long-term consequences, such as sleep related irregularities and/or cardiovascular diseases. This paper proposes a multimodal and feature selection-based processing pipeline to detect OSA as a computer-based alternative way to clinical polysomnography (PSG) method. In the proposed method, the oxygen saturation (SpO2) and the electrocardiogram (ECG) signals are fused at the feature-level for the classification. Five feature selection methods, namely Relieff, Chi-Square, Information Gain (IG), Principal Component Analysis (PCA), and Gain Ratio (GR) were applied to the problem to obtain robust features from both signal sources and to reduce the feature dimensionality. The effectiveness of utilized feature selection methods was analyzed using the Support Vector Machine (SVM), k-nearest neighbor (k-NN), and Naive Bayes (NB) classifiers. The experimental results on the real clinical samples from the PhysioNet dataset show that the proposed multimodal and feature selection-based method improves the classification accuracy, significantly.

***Keywords*—** feature fusion, sleep apnea recognition, electrocardiogram (ECG), saturation of oxygen (SpO2), SVM, NB, k-NN

Obstrüktif Uyku Apnesi Tanıma için Öznitelik Seçimi

***Özet*—** Obstrüktif uyku apnesi (OUA), uyku sırasında anormal nefes durması veya azalması ile sıkça tanımlanan yaygın bir uyku bozukluğudur. Bu hastalık, uyku ile ilgili düzensizlikler ya da kardiyovasküler hastalıklar gibi uzun vadeli sonuçlara yol açabilir. Bu çalışmada, uyku apnesi tanıma için, klinik polisomnografi (PSG) yöntemine alternatif olarak, çok kipli öznitelik kullanımı ve seçimine dayalı sayısal bir yöntem önerilmektedir. Önerilen yöntem, elektrokardiyogram (EKG) ve oksijen doygunluğu (SpO2) olarak adlandırılan iki fizyolojik sinyalin öznitelik düzeyli kaynaşımına dayalıdır. Her iki sinyal kaynağından da sağlam özellikler elde etmek ve öznitelik boyutunu azaltmak için Relieff, Chi-Square, Bilgi Kazancı (BK), Temel Bileşen Analizi (TBA) ve Kazanç Oranı (KO) olmak üzere beş öznitelik seçim yöntemi probleme uygulanmıştır. Elde edilen çok kipli öznitelikler ile Naive Bayes (NB), en yakın komşu (kNN) ve Destek Vektör Makinesi (DVM) sınıflandırıcıları tasarlanmış ve etkinlikleri sınanmıştır. PhysioNet veritabanındaki gerçek örnekler üzerinde yapılan deneysel çalışmalar, önerilen yöntemin sınıflandırma başarımını artırdığını göstermektedir.

***Anahtar Kelimeler*—** öznitelik kaynaştırma, obstrüktif uyku apnesi tanıma, EKG, SpO2, DVM, NB, k-NN

1. INTRODUCTION

As a type of sleep disorder, the obstructive sleep apnea (OSA) can be described by the obstruction of breathing during the sleep and recurring of this event. Particularly, a patient cannot breathe for seconds (typically 10-30 seconds) and this situation can occur many times in one night [1]. The frequent occurrence of OSA may cause several symptoms such as sleepiness, memory deficits, and depression. Undiagnosed OSA is also associated with diseases such as arrhythmias, brain stroke, ischemia, and higher blood pressures [2] and this results in billions of dollars costs due to the sleepiness-related accidents and health costs to the governments.

According to [3], the amount of OSA diagnoses in US is reported as approximately 6.62% of the population. The OSA not only causes sleep disorder, yet it may form the basis of other diseases such as diabetes, obesity, hypertension, and depression [4]. Hence, it clearly reduces the life-quality and -productivity and increases mortality. Reports also show that approximately 18 million people are formally diagnosed as OSA patients, while approximately 63 million are not diagnosed, yet. Due to this great negative impact of OSA in human healthcare, considerable research has been conducted on the recognition and monitoring of OSA in the literature. The research on OSA can be mainly divided into two categories, which are referred to as clinical (e.g., Polysomnography - PSG) and computational methods. The PSG test is an extensive test, which employs several sensors that are attached to the patient's body and being performed in sleep laboratory settings. On the other hand, the computational methods typically make use of sensor data that are collected from patients during the OSA event [5]-[8]. Typical sensor data includes oxygen saturation (SpO2), respiration rate, and heart rate signals.

The OSA is evaluated with the apnea-hypopnea index (AHI) and defines the number of OSA events per hour (e/h). A person with AHI index greater or equal to 15 is called as an OSA patient. Physiologically, this also means a person stops breathing more than 10 seconds and desaturation of oxygen in the blood. Typical treatments include surgery based (e.g., upper airway), therapy-based (e.g., CPAP [9]), and losing weight methods. The PSG is the most widely used and reliable standard for the OSA diagnosis, which records electrooculogram (EOG), respiratory movements, electroencephalogram (EEG), electrocardiogram (ECG), oxygen saturation, breath airflow, electromyogram (EMG), and body position. However, it has some practical drawbacks such as whole night measurement with several cables connected to the body and under the supervision of the medical officials in laboratory settings [10]. Also, the PSG is an expensive test with limited availability [11] and therefore receives some criticisms from researchers. As a result, simpler and reliable solutions are needed for OSA diagnosis.

It has been stated that the heart-rate intervals of the ECG signal are associated with the events during sleep apnea by means of some patterns. That is, during sleep apnea event, the ECG signal includes bradycardia followed by tachycardia at the end. Lavie successfully utilized this information to diagnose patients as sleep apnea [5]. Some other studies also showed the heart-rate variability (HRV) intervals (RR) can be used to recognize sleep apnea [12]. Chazal et al. also use the statistical representations of signals extracted from HRV and ECG-derived respiratory signal (EDR) [11]. Some attempts make use of single-lead ECG to classify sleep apnea [13]. Yilmaz et al., use RR statistical representations of RR intervals as the features and design classifiers using the Support Vector Machines (SVM), Quadratic Discriminant Analysis (QDA), and k-nearest neighbor (k-NN) [14]. Espinoza-Cuadros et al. analysis the machine learning algorithms and features sources for the detection of OSA [15]. Some important conclusions include feature sources (e.g., speech) may be highly correlated with the subject characteristics such as gender and age. Secondly, as the number of features (or feature dimensions) increases compared to the number of patients, overfitting may occur. Therefore, feature selection approaches can solve this problem by selecting the proper subset from high-dimensional feature spaces.

To the best of our knowledge, current studies on computational OSA recognition mostly tackle with single modal approaches, e.g., make use of parameters from single source modality. There also exist some studies using multimodal signal sources for the recognition but do not consider feature selection [16]. In this study, the effect of different feature selection methods on the recognition accuracy of OSA classification is analyzed in the multimodal feature fusion context. To this end, the SpO2 and the ECG physiological signals are fused at the feature level and extensive tests regarding individual- and joint-features performances are presented using five feature selection methods with three classifiers. The proposed method is also referred to as MM-PCA-SVM throughout the paper.

The structure of the paper is as follows. Related studies are presented in Section 2. In Section 3, we introduce our methodology and describe the proposed method. The results and evaluations are presented in Section 4. Finally, we conclude the study in Section 5.

2. RELATED WORK

The ECG signal has a common usage for the detection of OSA due to its association with the OSA events. Therefore, time- and frequency-domain features are extracted from this signal and fed to machine learning (ML) algorithms for further classification. One attempt towards this direction is the study of Khandoker et al., in which they extract RR intervals obtained from ECGs and respiratory (EDR) signals by wavelet transformation and trained SVM classifier for the detection. In the study around 90% of the subjects has been correctly classified as OSA [17]. In some studies, instead of classifying the whole ECG signal, the ECG signals are analyzed in segment-basis and the classification is performed for each of these segments. In this case, the time- and frequency-domain features are extracted from each fixed-size segment. This type of methods detect the existence of OSA for each analyzed segment and can be used to determine the severity of OSA [18]-[19]. Chazal et al. make use of solely the ECG signal for the detection of OSA. They focus on different representations of the ECG signal such as statistical values of RR intervals and respiratory signals and use linear discriminant (LD) classifier for the detection [20]. They show that the classification accuracy of around 85% is achievable by using the RR intervals.

Although the OSA prevalence is mostly associated with adult people, can also occur among the children. Shouldice et al. investigate this situation for pediatric subjects. To this end, they perform quadratic discriminant analysis (QDA) by using the 1-minute segmented ECG signals as the feature and report an accuracy of 72.1% for segment-based evaluations, where subject-base accuracy is 84% [18]. Mendez et al. also perform segment-based analyses on the ECG signals and train a neural network (NN) and kNN classifiers. They achieve 88% classification accuracy [21]. Another study addresses the real-time detection of OSA events and develops a smart-phone based sleep apnea monitoring system [22]. They use 111-dimensional feature (63 elements from RR and 48 from the EDR) to train the SVM classifier.

Considerable research focuses on the simple and convenient-way of OSA diagnosis. Some of them make use of bio-signal sensors such as nasal airflow, Oximeter, ECG, and EEG [6]. The oxygen level in the blood (SpO2) is used for OSA screening [23], [24]. Yen et al. address the non-invasive ways of OSA and SA detection [25]. They utilize CPAP signal for identifying OSA events. Patangay et al. study the detection of OSA in heart disease subjects. They combine the ECG signal with heart sound, which is recorded during OSA events [26].

In [27], classifier performances are evaluated for the classification of OSA syndrome. They examine 58 positive and 25 negative OSA patients using four classifiers of which C4.5 demonstrates the best performance. Chen et al. [28] propose a feature-selection method and apply the method to the diagnosis of OSA. In their experiments, the Particle Swarm Optimization (PSO) based method along with 1-NN classifier improves prediction performance compared with back propagation neural networks, logistic regression, SVM, and C4.5. However, the methods efficiency is bounded by the exploration and exploitation problems of PSO. Espinoza-Cuadros et al. explore the relation between the speech signal and the OSA [15]. They use spectral features obtained by i-vectors and train support vector regression (SVR) for the prediction. They obtain best results with i-vectors and SVR with linear kernel. Blanco et al. introduce a novel method to detect OSA based on patient's voices [29] on the manually collected dataset. In [30], Aydogan et al. performs visual scoring of 74 patients using two methods, namely morphological filter and ANN-based to diagnose OSA. They analyze the scoring success of both methods and obtain an average accuracy of 88.33% and 87.28% for ANN and morphological filter-based methods, respectively. Ucar et al. aim to detect respiratory arrests in OSA patients [31]. They use only one signal, the PPG, and extract 34 features for the analyses. They conclude that PPG signals and the respiratory arrests have connection and the PPG signal can be used for the diagnosis of OSA.

One attempt to the multimodal OSA detection is made by Rutkowski [16]. In this study, multimodal biomedical signals (brainwaves and peripheral physiological) are used for identifying 9 sleep apnea episodes via a data-driven approach. Although the proposed automatic apnea detection method can be used as a solution for sleep obstruction suffering patients, this may require special setups due to the required nine biomedical signal measurements to make such a detection. A more lightweight design is presented in [32]. Here, Lee et al. develop a real-time system for the identification of hypopnea and sleep apnea events. They use a nasal pressure signal that are collected from a manually collected dataset of fifty patients. Though the system provides SA screening, the method is highly correlated with the rules that are defined on the collected dataset. Kim et al. explore the problem of severity classification of OSA using the patient breathing-sound [33]. They extract sound energy and statistical properties derived from spectral density and build SVM classifier to detect OSA severity. Using the method on the SNUBH breathing sound dataset, they report classification accuracies of 92.78% and 79.52% for the subject and severity classification, respectively. However, sounds signals are highly sensitive to the recording/patient characteristics and environment conditions, which may require additional processing steps. Recently, deep NN models achieve state-of-the art results in several application domains. A recent attempt on OSA detection using the Convolutional Neural Networks (CNNs) is presented by Dey et al. [34]. They use a single-modal approach in which only using a single-lead ECG signal is used for training the CNN. They report an accuracy of 98.91% on the Physionet dataset when the dataset is separated into two folds for training and testing. However, the study considers only the absence or presence of apnea and does not consider the apnea severity.

In [13], Vimala et al. introduce a system for SA classification based on the electroencephalogram (EEG) signal. They employ the SVM, kNN, and the ANN methods for the classification. Their results on the privately collected EEG signals from 18 sleep apnea patient show that the SVM classifier performs better than the others. Urtnasan et al. use the ECG feature (single-lead) with a convolutional neural network (CNN) for the detection of OSA [19]. The results are presented on a dataset including full-night PSG data collected from 82 subjects in total (63 of them is male-record and 19 is female-record) diagnosed as OSA. They indicate that the single-lead ECG signal learned by a CNN model performs better than the hand-crafted features from the same signal. However, the method does not classify OSA severity due to the utilized dataset and there is no cross-checking of the collected reference PSG data since the scores are performed by one certified clinician. One attempt that focuses on the feature representation make use of wavelet coefficients extracted from the single-lead ECG [2]. In the study, the best features are selected by the sequential feature selection (SFS) method and used with various classifiers, such as SVM, LDA, ANN, kNN, and NB. They achieve best accuracy of 91.81% using the RBF SVM on the Physionet dataset. Also, there exist some studies employing facial features other than the physiological features. Islam et al. [35], investigate the depth map of human facial scans for OSA prediction. They make use of the 2D depth maps obtained from 3D facial scans of patients and perform transfer learning. The patient prediction results for the moderate OSA severity ($15/h<AHI\leq30/h$) is promising with

accuracies around 60% but need further improvement since the utilized dataset is very small.

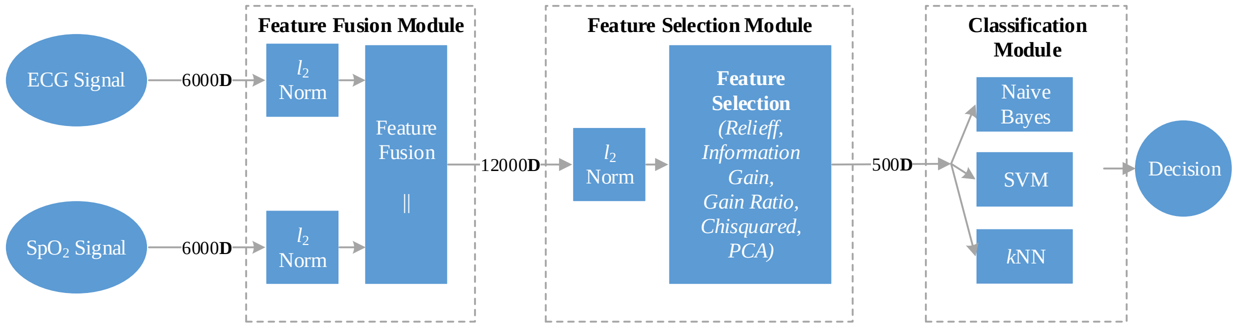

Figure 1. Illustration of the proposed processing pipeline for sleep apnea recognition

One of the main problems in OSA detection studies is the high-dimensional characteristics of the utilized features. This situation mainly attributed by data redundancy and may cause to reduce the learner performance. Different feature selection schemes in the literature including principal component analysis [36], wrapper methods [21], and statistical evaluation methods [22] have been used for dimension reduction or robust feature representation purposed in several computer vision (CV) tasks, but have not been considered in sleep apnea classification as much as in CV.

3. PROPOSED METHOD

The entire system can be broken down into three disjoint modules: multimodal feature fusion, feature selection, and sleep apnea classification. We show the general structure of the proposed system in Figure 1.

3.1. Multimodal Approach and Feature Fusion

In the multimodal analysis, multiple sources of information are used to achieve a specific goal, whereas single modal analysis makes use of only one signal source. Multimodal approaches demonstrate good performances in various computer vision tasks such as video object/event/action detection and classification and in situations where multiple modalities (e.g., visual, audio, text) exist. Fusing different information collected from different signal channels may provide robustness towards varying system dynamics, and hence may increase recognition accuracy. By this motivation and due to the close association of these signals with OSA, we make use of the SpO2 and ECG signals in the multimodality context.

Information fusion can be performed at different levels such as data or decision fusion. Here we consider the fusion at the feature level by combining the SpO2 and ECG signals. Thus, we aim to capture as much information as possible carried by each modality. There are various methods (e.g., sum/max/average pooling, concatenation) in the literature for combining the information in each modality. We employ the concatenation operator due to its properties, which are simplicity, preserving as much information as possible for further feature selection, and to keep temporal information within the signals. Let $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_m)$ are the feature vectors representing different information sources and *n* and *m* corresponds to the vector dimensions, respectively. The concatenated feature vector is denoted by F_s and derived from $X \parallel Y$ where ‖ represents the concatenation operator:

$$F_s = \langle x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m \rangle \quad (1)$$

Since the concatenation operator preserves all the information from different information sources, it increases the feature dimension with respect to the combined feature dimensions and may arise curse of dimensionality problem for the learner. Use of feature selection method is a good way to cope with this problem, yet to provide more robust and representative features.

3.2. Feature Selection

Feature selection is an important step for pattern recognition tasks, since it improves the performance of classification algorithms and may reduce the complexity of the computations by removing redundant and/or unnecessary information. Feature selection is also widely applied when different modalities of information sources are existing and we do not have a prior knowledge regarding how to combine them [37]. In our study, we employed five feature selection algorithms, namely Gain Ratio (GR), Chi-Square ($\mathcal{X}^2$), Information Gain (IG), Principal Component Analysis (PCA), and Relieff [38]-[39] to obtain robust feature subsets from the ECG and SpO2 signals. Although there have been several types of algorithms for this purpose, we prefer to use filter-based methods due to their simplicity and classifier independence.

3.2.1 Information Gain (IG)

Information Gain (IG) is a statistical property that measures the separation ability of each training instance attribute for desired target classification. To this end, the method evaluates the gain of each feature by using the entropy (S) measure. Simply, the entropy states the number of bits of information needed to encode the classification of an arbitrary instance of the dataset, D. If the target classification is multi-class and we have c categories, the entropy of D relative to c is calculated as

$$Entropy(D) = S = \sum_{i=1}^{c} -p_i log_2 p_i \tag{2}$$

where p_i is the proportion of D for class i [40]. Let an instance of the D is described by the attribute set of $A = (a_1, a_2, \dots, a_n)$, where n is the number of attributes of an instance in D. Hence, the information gain of attribute a_j in A relative to D is

$$IG(D, a_j) = Entropy(D) - \sum_{v \in Values(a_j)} \frac{|D_v|}{|D|} Entropy(D_v) \tag{3}$$

where $D_v = \{d \in D | a_j(d) = v\}$ and $Values(a_j)$ represents the set of all possible values for attribute a_j.

3.2.2 Gain Ratio (GR)

The number of possible values of an attribute ($Values(a_j)$) is crucial for IG since an attribute having more possible values dominate the attribute that have fewer possible values over the dataset. This leads to higher IG for bigger $Values(a_j)$ relative to the dataset, although the attribute with more values has a very poor classification performance. In order to eliminate this issue, we also consider the gain ratio [39] method, which adjusts the IG by using split information as follows:

$$GR(D, a_j) = \frac{IG(D, a_j)}{-\sum_{i=1}^{m} \frac{|D_i|}{|D|} log_2 \frac{|D_i|}{|D|}} \tag{4}$$

where *m* represents the number of values of attribute a_j.

3.2.3 Chi-Square

In general, the chi-square feature selection method measures the independence of two events. In our case, these two events are referred to as the feature and its class over the dataset:

$$\mathcal{X}^2 = \sum_{k \in \{0,1\}} \sum_{l \in \{0,1\}} \frac{(N_{kl} - E_{kl})^2}{E_{kl}} \tag{5}$$

where *k* and *l* represent the existence (0 or 1) of feature and its class, respectively. Here, N_{kl} is the observed frequency in the dataset and E_{kl} is the expected frequency. We calculate a score for each individual feature using the method (5) over the dataset and select a predefined size for the final feature set.

3.2.4 Relieff

The Relieff algorithm relies on the basic idea of estimating the quality of attributes by searching nearest neighbors of randomly selected instances based on their values [41]. The algorithm is capable of dealing with incomplete and noisy data. In order to find the nearest-neighbors, the algorithm calculates the distance between the instances and treats the missing values of attributes probabilistically. For instance, the difference between an attribute value a_j in $\boldsymbol{A}$ for two example instances d_i and d_k ($\{d_i, d_k \in D\}$) is calculated as

$$Difference(a_j, d_i, d_k) = 1 - P(value(a_j, d_k) | class(d_i)) \tag{6}$$

where $P(.)$ denotes prior probability. If both instances have unknown value:

$$Difference(a_j, d_i, d_k) = 1 - \sum_{v \in Values(a_j)} \Big(P\big(v|class(d_i)\big) \times P\big(v|class(d_k)\big)\Big) \tag{7}$$

where $Values(a_j)$ represents the set of all possible values for attribute a_j. In our evaluations, we empirically select the first 500 features from the feature selection methods based on the experiments.

3.2.5 Principal component analysis (PCA)

Basically, the PCA method is a type of dimension-reduction method and performs mathematical procedures to transform a number of (likely) correlated variables to some number of uncorrelated variables called principal components [42]. It aims to preserve the linear structure intact when transforming into a low-dimensional space. Since we use different information sources (ECG and SpO2) and the PCA concerns with the variance maximization, we perform ℓ_2 normalization prior to applying the PCA.

3.3. Classifier Design

This section presents the ML algorithms we employed for the recognition of OSA. We first describe our design considerations of the k-nearest neighbor (kNN), Naive Bayes (NB), and Support Vector Machine (SVM) algorithms and then present performance comparisons among them.

The NB classifier is a statistical method based on the Bayesian rule [43]. We implement the NB algorithm based on the Gaussian distribution in Matlab Statistics and

Machine Learning Toolbox [16]. In this method, the mean and covariance matrix are computed using the training data and in the testing phase, conditional probabilities of the categories are calculated for a given pattern and subsequently, the posterior probability is computed. The test pattern is classified into a specific category using the posterior probability, for which the posterior probability is highest. The kNN algoritm is a type of non-parametric method, which uses the local neighborhoods to calculate the prediction. The algorithm simply stores the position of training samples and their categories and classifies new instances using a similarity measure. The category of a new instance is decided based on majority voting of its k-neighbors. The similarity measure and the value of k are set to *Euclidean* distance and 3, respectively based on the experiments on parameter k. The third algorithm we employed for the classification is SVM due to its success in medical diagnosis tasks [44]-[45]. The Libsvm package has been used for the implementation of the SVM [46]. The SVM kernel has been set to the radial basis function (RBF) and its C and γ parameters are optimized using the grid search algorithm.

In the following section, we demonstrate the effectiveness of the proposed approach using different test scenarios. To this aim, we consider subject- and severity-independent situations. We present the results using the designed classifiers on a performance dataset in terms of accuracy.

4. EXPERIMENTS AND EVALUATIONS

4.1. Dataset

We perform the evaluations on the Physionet dataset, which consists of SpO2 and ECG signals [47]. The database includes 70 records of which half of it is reserved for training and the rest is for testing. The length of each record is 7-10 hours and consists of continuous ECG signal, apnea annotation, and QRS annotations. The apnea annotations are performed by human experts based on the recorded respiration signal, simultaneously. The records with number a01-a04, b01, and c01-c03 includes additional signals, which are Resp A, Resp C, chest and abdominal respiratory effort signals, and Resp N. The records in the dataset are separated into three groups, which are Group A, Group B, and Group C. Group A includes the whole night sleeps of 20 subjects with high severity of apnea, Group B includes the whole night sleeps of 5 subjects having low severity of apnea, and lastly, Group C contains the whole night sleeps of 10 subjects without apnea. The annotations are performed by human experts as absence and presence of apnea for each 1-minute segment [18].

The dataset uses the AHI index for diagnosis, which provides the indication of OSA severity. This index is calculated based on the number of OSA events per hour. As a standard, the AHI index classifies the OSA severity (OSA_s) into three categories as follows:

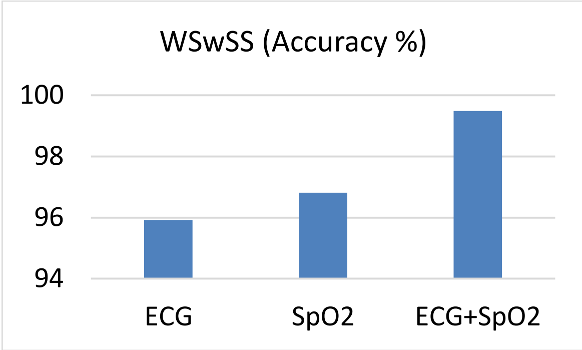

Figure 2. Single- and multi-modal feature performances for the WSwSS scenario.

$$OSA_{s.} = \begin{cases} mild \quad , & 5/h \leq AHI \leq 15/h \\ moderate, & 15/h < AHI \leq 30/h \\ severe \quad , & AHI > 30/h \end{cases} \qquad (8)$$

Each 1-minute record has been also annotated as apnea (A) or not apnea (normal) in the dataset by the sleep experts based on the collected measurements (e.g., ECG, respiration, SpO2, airflow, etc.).

4.2. Results and Evaluations

Most of the previous studies perform the tests within each SA severity (e.g., mild, moderate, and severe). That is, the train and tests sets include the instances of the same severity and does not include the severities from different categories. We name this situation as severity-dependent tests. We use three test scenarios [49] to evaluate the effectiveness of our method:

- Within Subject with Same Severity (WSwSS)
- Between Subjects with Same Severity (BSwSS)
- Between Subjects with Different Severity (BSwDS)

For the WSwSS scenario, we use a subset of the dataset including the same subject's same severity records. We use this subset for both in training and testing phases. The BSwSS scenario includes the records from all subjects with same severities, i.e., all the records within this subset have the same severity. Although prior two scenarios are common in the literature, the BSwDS scenario is new to our study and considered to evaluate the effectiveness of the methods since this scenario includes all subjects having distinct OSA severities in testing and training phases. To this aim, we use all the records of subjects with different severities, which are named as A01, A02, A03, A04; B01; C01, C02, and C03 in the dataset. In all scenarios, we never use an instance in testing if we used it in training and vice versa. In order to be comparable with similar studies in [2], [48],[27], 10-fold cross-validation is used for the WSwSS scenario. For the BSwSS, we keep one subject (e.g., A03) for testing and use the rest (e.g., A01, A02, and A04) for training purpose. Similarly, for the BSwDS, we keep one subject (e.g., C02) for testing and use the rest of the dataset (e.g., A01, A02, A03, A04; B01; C01, and C03) for the training. For the other two scenarios, we repeat this procedure by excluding each subject for testing phase and using the rest for training.

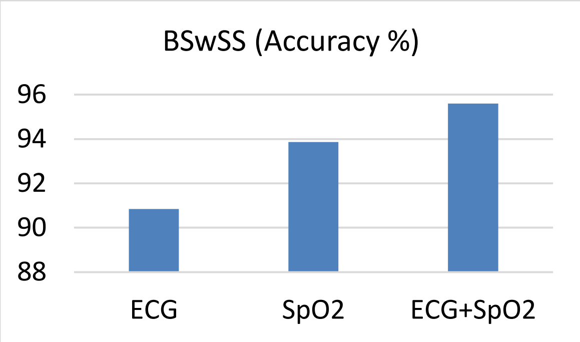

Figure 3. Single- and multi-modal feature performances for the BSwSS scenario.

The results of the WSwSS scenario is presented in Figure 2. We repeat the test for each subject and give the result as average accuracy. For instance, we perform the test for the records of subject A01 with the same severity using the 10-fold cross-validation and calculate the accuracy. We repeat this procedure for each subject and calculate the average accuracy in the end. According to the results, the best classification performance is achieved by the multimodal approach using the RBF-SVM with an average accuracy of 99.49%. The individual signals SpO2 (96.81%) and ECG (95.92%) follow this performance in order.

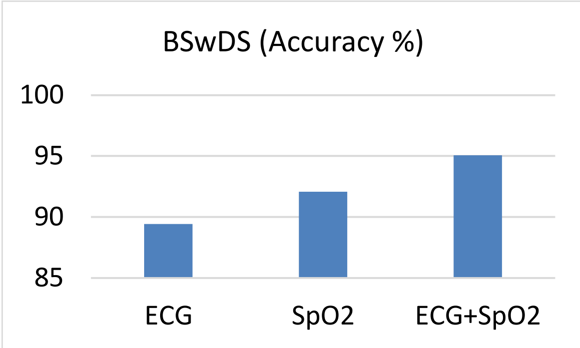

Figure 4. Single- and multi-modal feature performances for the BSwDS scenario.

Table 1. Overall classification results. Bold numbers indicate the best results.

Feature \ Test Scenario / Feature Selection Method	Multi-modal (ECG+S_pO_2)								
	WSwSS			BSwSS			BSwDS		
	RBF SVM	kNN	NB	RBF SVM	kNN	NB	RBF SVM	kNN	NB
Without Feature Selection	98,50	87,07	82,63	79,69	81,08	82,22	86,16	75,21	78,78
Chi-Square (Chi)	99,49	79,73	94,78	79,69	81,08	82,27	86,20	81,59	78,80
Information Gain (IG)	99,49	80,59	94,74	85,15	85,22	76,95	86,26	81,77	78,59
Gain Ratio (GR)	96,62	81,59	95,32	79,44	81,34	80,58	86,10	81,61	78,67
Relieff	98,50	82,63	87,07	82,10	82,22	82,83	86,16	78,78	81,47
PCA	**99,49**	97,65	92,19	**95,60**	80,96	88,86	**95,07**	81,47	89,13
Single-modal (ECG)									
Without Feature Selection	89,43	86,81	87,25	88,47	78,95	88,17	88,86	84,35	73,28
Chi-Square (Chi)	89,44	87,37	87,48	88,47	78,77	88,19	89,43	84,47	78,35
Information Gain (IG)	89,44	87,99	88,42	88,49	78,89	88,26	89,43	84,32	78,12
Gain Ratio (GR)	89,44	87,68	95,32	88,47	78,67	80,58	89,34	84,26	78,67
Relieff	89,43	87,25	86,81	89,43	88,72	81,09	89,43	73,28	84,35
PCA	**95,92**	94,23	95,64	**90,85**	88,18	88,40	**89,43**	89,38	85,10
Single-modal (S_pO_2)									
Without Feature Selection	95,46	94,57	80,71	90,41	92,05	81,28	89,03	88,70	83,12
Chi-Square (Chi)	95,56	94,63	79,03	91,67	81,21	80,72	89,20	78,45	80,66
Information Gain (IG)	95,38	94,67	79,94	91,69	82,90	80,66	88,62	79,32	81,26
Gain Ratio (GR)	95,56	94,96	95,32	92,88	84,75	80,58	90,58	81,88	78,67
Relieff	95,46	80,71	94,57	93,75	81,28	92,53	**92,08**	83,12	90,70
PCA	**96,81**	91,13	87,09	**93,87**	93,52	85,17	90,08	90,95	81,91

The results of the BSwSS scenario is depicted in Figure 3. Results show that the proposed feature selection-based method yields the best performance with an accuracy of 95.60%. In terms of individual features, the SpO2 signal performs better than the ECG signal. We observe that performance difference between the individual signals in BSwSS is bigger than the difference in WSwSS. This may show the SpO2 signal is better in capturing the subject changes than the ECG signal or can be related to dataset characteristics.

tests, we set the classifier parameters empirically (RBF kernel w/SVM and k=3 for the kNN). The kNN algorithm gives competitive classification performance compared to the SVM when using the ECG features. We observe that when using the RBF-SVM, while the ECG signal performs reasonable performance for the WSwSS scenario, its performance degrades in the BSwDS and BSwSS tests. In all experiments, we achieve better recognition rates with the RBF-SVM than the NB and kNN.

Table 2. Comparison of our method with the literature

Method	**WSwSS** (Acc %)	**BSwSS** (Acc %)	**BSwDS** (Acc %)
Shi et al. [48]	97.3	89.8	-
SVM [27]	82.24	-	-
C4.5 [27]	80.91	-	-
Bagging Reptree [27]	84.4	-	-
Bagging ADtree [27]	79.85	-	-
FT Trees [27]	79.32	-	-
AdaBoost [27]	77.79	-	-
REP Tree [27]	81.33	-	-
kNN [27]	81.65	-	-
Decision Table [27]	80.79	-	-
MLP [27]	81.6	-	-
Our Method (MM-PCA-SVM)	**99.49**	**95.6**	**95.07**
kNN [2]	92.52	-	-
ANN [2]	91.36	-	-
LDA [2]	79.92	-	-
QDA [2]	61.16	-	-
GentleBoost [2]	89.27	-	-
NB [2]	44.04	-	-
LR [2]	73.7	-	-
SVM(RBF) [2]	92.98	-	-
SVM(POL) [2]	90.3	-	-
SVM(LIN) [2]	82.65	-	-

Figure 4 shows the results of the BSwDS scenario. Again, the proposed method performs better than the individual feature performances with an accuracy of 95.07%. The individual features perform very competitive results in this scenario and we do not observe a significant difference among the individual features.

To sum up, the proposed multimodal and feature selection-based method with RBF-SVM performs better in all tests compared to the individual features. The improvement of the proposed method for all scenarios is 3.64% on average (the difference between the best multimodal and individual feature performances in three scenarios, Table 1). On the other hand, the accuracy of the proposed method slightly decreases among the test scenarios WSwSS, BSwSS, BSwDS in order as an expected outcome of the research.

We compare classifier performances in Table 1. In the table, we also include the multi- and single-modal feature performances with the feature selection methods. In the

We compare the result of our best performing multimodal and feature selection based SVM method (MM-PCA-SVM) with the related works [2], [48], and [27] since they also employ either multimodal features or feature selection. Shi et al. make use of three sensor data (SpO2, ECG, and Breath) to perform OSA classification [48]. They also use SVM classifier with the features, which are extracted from each modality using the signal processing techniques. Our method outperforms Shi et al. for the tests of BSwSS and the WSwSS scenarios (Table 2). We cannot provide a comparison for the BSwDS scenario since its new to our study and is not evaluated in [48]. Xie et al. develop a real-time sleep apnea system utilizing the SpO2 and the ECG signals in the late-fusion scheme [27]. They experiment with various ML algorithms and achieve 84.40% classification accuracy using the Bagging.REPTree. Compared with our multimodal method, which performs better, the success may lie on the early-fusion scheme.

5. CONCLUSION

In this study, we present a multimodal and feature selection-based approach for computer-based classification of OSA. To this end, we extracted information from two physiological signals and combined them at the feature level. Five feature selection methods are applied to the problem in order to obtain robust and memory-efficient features. These features are then fed to selected ML algorithm for the classification task.

Our outcomes on the performance dataset clearly show that multimodal approach yields superior performance compared with the individual features. The SpO2 signal is better than the ECG signal in capturing the subject changes. The RBF-SVM gives the best results compared with the kNN and the NB methods. Experimented methods perform better for the WSwSS scenario while they degrade for the BSwDS scenario as expected. We believe that subject- and severity-independent tests can be used to better demonstrate the recognition ability of OSA. Although the feature selection methods make reduction in feature dimensions, the single-modal approaches may be preferred for personal usage to save the feature dimensionality introduced by the concatenation operator.

Future works may lie in two directions. The first one is to investigate more robust feature representations (e.g., deep autoencoders) and the second direction is to utilize deep models for the classification.

ACKNOWLEDGMENTS

The author thanks Dr. Gökhan Memiş for running the experiments.

REFERENCES

[1] W. W. Flemons, D. Buysse, S. Redline, A. Oack, K. Strohl, J. Wheatley, T. Young, N. Douglas, P. Levy, W. McNicolas, J. Fleetham, D. White, W. Schmidt-Nowarra, D. Carley, J. Romaniuk, "Sleep-related Breathing Disorders in Adults: Recommendations for Syndrome Definition and Measurement Techniques in Clinical Research", *Sleep*, 22(5), 667-689, 1999.

[2] A. Zarei, B. M. Asl, "Automatic Detection of Obstructive Sleep Apnea Using Wavelet Transform and Entropy-Based Features From Single-Lead ECG Signal", *in IEEE Journal of Biomedical and Health Informatics*, 23(3), 1011-1021, 2019.

[3] B. M. Altevogt, H. R. Colten, **Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem**, Institute of Medicine (US) Committee on Sleep Medicine and Research, Washington (DC): National Academies Press (US), 2006.

[4] D. Leger, "The cost of sleep-related accidents: a report for the National Commission on Sleep Disorders Research", *Sleep*, 17(1), 84-93, 2994.

[5] J. N. McNames, A. M. Fraser, "Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram", *Computers in Cardiology*, 27, 749-752, 2000.

[6] F. Mendonca, S. S. Mostafa, A. G. Ravelo-Garca, F. Morgado-Dias, T. Penzel, "Review of Obstructive Sleep Apnea Detection Approaches," *in IEEE Journal of Biomedical and Health Informatics*, 23(2), 825-837, 2019.

[7] L. Lavie, "Obstructive sleep apnoea syndrome an oxidative stress disorder," *Sleep Medicine Reviews*, 7(1), 35-51, 2003.

[8] M.O. Mendez, S. Cerutti, A.M. Bianchi, J. Corthout, S. Van Huffel, M. Matteucci, T. Penzel, "Automatic Screening of Obstructive Sleep Apnea from the ECG Based on Empirical Mode Decomposition and Wavelet Analysis", *Physiological Measurement*, 31(3), 273-289, 2010.

[9] R.K. Kakkar, R.B. Berry, "Positive Airway Pressure Treatment for Obstructive Sleep Apnea", *Chest.*, 132(3), 1057-1072, 2007.

[10] C. Armon, K. G. Johnson, A. Roy, W. J. Nowack, "Polysomnography", 2016.

[11] P. Chazal, T. Penzel, C. Heneghan, "Automated Detection of Obstructive Sleep Apnoea at Different Time Scales Using the Electrocardiogram", *Physiological Measurement*, 25, 967-983, 2004.

[12] T. Penzel, J. McNames, P. de Chazal, B. Raymond, A. Murray, G. Moody, "Systematic Comparison of Different Algorithms for Apnea Detection Based on Electrocardiogram Recordings", *Medical and Biological Engr. and Comp.*, 40(4), 402-407, 2002.

[13] V. Vimala, K. Ramar, M. Ettappan, "An Intelligent Sleep Apnea Classification System Based on EEG Signals", *Journal of medical systems*, 43(2), 36, 2019, https://doi.org/10.1007/s10916-018-1146-8.

[14] B. Yilmaz, M. H. Asyali, E. Arikan, S. Yetkin, F. Ozgen, "Sleep Stage and Obstructive Apneaic Epoch Classification using Single-lead ECG", *BioMedical Engineering*, 9(1), 39, 2010.

[15] F. Espinoza-Cuadros, R. Fernndez-Pozo, D. T. Toledano et al., "Reviewing the connection between speech and obstructive sleep apnea", *BioMedical Engineering*, 15, 20, 2016.

[16] T.M. Rutkowski, "Data Driven Multimodal Sleep Apnea Events Detection", *Journal of Medical Systems*, 40, 162, 2016.

[17] A. H. Khandoker, M. Palaniswami, C. K. Karmakar, "Support Vector Machines for Automated Recognition of Obstructive Sleep Apnea Syndrome From ECG Recordings", *in IEEE Transactions on Information Tech. in Biomedicine*, 13(1), 37-48, 2009.

[18] R.B. Shouldice, L.M. O'Brien, C. O'Brien, P. de Chazal, D. Gozal, C. Heneghan, "Detection of Obstructive Sleep Apnea in Pediatric Subjects using Surface Lead Electrocardiogram Features", *Sleep*, 27(4), 784-792, 2004.

[19] E. Urtnasan, J. U. Park, E. Y. Joo, K. J. Lee, "Automated detection of obstructive sleep apnea events from a single-lead electrocardiogram using a convolutional neural network", *Journal of medical systems*, 42(6), 104, 2018, https://doi.org/10.1007/s10916-018-0963-0.

[20] P. de Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, M. O'Malley, "Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea", *in IEEE Tran. on Biomedical Engineering*, 50(6), 686-696, 2003.

[21] M. O. Mendez, A. M. Bianchi, M. Matteucci, S. Cerutti, T. Penzel, "Sleep Apnea Screening by Autoregressive Models From a Single ECG Lead", *in IEEE Transactions on Biomedical Engineering*, 56(12), 2838-2850, 2009.

[22] M. Bsoul, H. Minn, L. Tamil, "Apnea MedAssist: Real-time Sleep Apnea Monitor Using Single-Lead ECG", *in IEEE Transactions on Information Technology in Biomedicine*, 15(3), 416-427, 2011.

[23] N. Oliver, F. Flores-Mangas, "HealthGear: a real-time wearable system for monitoring and analyzing physiological signals", **InternationalWorkshop onWearable and Implantable Body Sensor Networks (BSN'06)**, Cambridge, MA, 4-64, 2006.

[24] B. Raymond, R. M. Cayton, R. A. Bates, M. Chappell, "Screening for obstructive sleep apnoea based on the electrocardiogram-the computers in cardiology challenge", *Computers in Cardiology 2000*, 27 (Cat. 00CH37163), 267-270, Cambridge, MA, 2000.

[25] Fu-Chung Yen, K. Behbehani, E. A. Lucas, J. R. Burk, J. R. Axe, "Noninvasive technique for detecting obstructive and central sleep apnea", *in IEEE Tran. on Biom. Engr.*, 44(12), 1262-1268, 1997.

[26] A. Patangay, P. Vemuri, A. Tewfik, "Monitoring of Obstructive Sleep Apnea in Heart Failure Patients", **2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society**, Lyon, 1043-1046, 2007.

[27] B. Xie, H. Minn, "Real-Time Sleep Apnea Detection by Classifier Combination", *in IEEE Transactions on Information Technology in Biomedicine*, 16(3), 469-477, 2012.

[28] L. F. Chen, C. T. Su, K. H. Chen et al, "Particle swarm optimization for feature selection with application in obstructive sleep apnea diagnosis", *Neural Computing and Applications*, 21(8), 2087-2096, 2012.

[29] J. L. Blanco, L. A. Hernndez, R. Fernndez et al., "Improving Automatic Detection of Obstructive Sleep Apnea Through Nonlinear Analysis of Sustained Speech", *Cognitive Computation*, 5(4), 458-472, 2015.

[30] O. Aydogan, A. Oter, K. Guney et al., "Automatic Diagnosis of Obstructive Sleep Apnea/Hypopnea Events Using Respiratory Signals", *Journal of Medical Systems*, 40, 274, 2016.

[31] M. K. Ucar, M. R. Bozkurt, C. Bilgin, K. Polat, "Automatic detection of respiratory arrests in OSA patients using PPG and machine learning techniques", *Neural Computing and Applications*, 28(10), 2931-2945, 2016.

[32] H. Lee, J. Park, H. Kim et al., "New Rule-Based Algorithm for Real-Time Detecting Sleep Apnea and Hypopnea Events Using a Nasal Pressure Signal", *Journal of Medical Systems*, 40, 282, 2016.

[33] J. Kim, T. Kim, D. Lee et al., "Exploiting temporal and nonstationary features in breathing sound analysis for multiple obstructive sleep apnea severity classification", *BioMedical Engineering*, 16, 6, 2017.

[34] D. Dey, S. Chaudhuri, S. Munshi, "Obstructive sleep apnoea detection using convolutional neural network based deep learning framework", *Biomedical engineering letters*, 8(1), 95-100, 2018, https://doi.org/10.1007/s13534-017-0055-y.

[35] S. M. Islam, H. Mahmood, A. A. Al-Jumaily, S. Claxton, "Deep Learning of Facial Depth Maps for Obstructive Sleep Apnea Prediction", **2018 International Conference on Machine Learning and Data Engineering (iCMLDE)**, Sydney, Australia, 154-157, 2018.

[36] S. M. Isa, M. I. Fanany, W. Jatmiko, A. M. Arymurthy, "Sleep Apnea Detection from ECG Signal: Analysis on Optimal Features, Principal Components, and Nonlinearity", **2011 5th International Conference on Bioinformatics and Biomedical Engineering**, Wuhan, 1-4, 2011.

[37] N. Xiong, P. Svensson, "Multi-sensor management for information fusion: issues and approaches", *Information Fusion*, 3(2), 163-186, 2002.

[38] I. Kononenko, E. Simec, M. R. Sikonja, "Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF", *Applied Intelligence*, 7(1), 39-55, 1997.

[39] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, 1(1), 81-106, 1986.

[40] Thomas M. Mitchell, **Machine Learning**, 1 ed., McGraw-Hill, Inc., New York, NY, USA, 1997.

[41] M. Robnik-Sikonja, I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF", *Machine Learning*, 53(1), 23-69, 2003.

[42] I. Jolliffe, "Principal Component Analysis", **International Encyclopedia of Statistical Science**, Editör: Lovric M., Springer, Berlin, Heidelberg , 2011.

[43] R. O. Duda, P. E. Hart, D. G. Stork, **Pattern Classification**, Wiley Interscience, 2000.

[44] M. H. Calp, "Medical Diagnosis with a Novel SVM-CoDOA Based Hybrid Approach", *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 9(4), 6-16, 2018.

[45] T. B. Alakuş, İ. Türkoğlu, "Pozitif ve Negatif Duyguların Ayrımında Etkili EEG Kanallarının Dalgacık Dönüşümü ve Destek Vektör Makineleri ile Belirlenmesi", *Bilişim Teknolojileri Dergisi*, 12(3), 229-237, 2019.

[46] C.C Chang, C.J. Lin, "LIBSVM: a library for support vector machines", *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 27, 2011.

[47] A. L. Goldberger, L. Amaral.et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals", *Circulation*, 101(23), e215-e220, 2000.

[48] C. Shi, M. Nourani, G. Gupta, L. Tamil, "Apnea MedAssist II: A smart phone based system for sleep apnea assessment", **2013 IEEE Intl. Conf. on Bioinformatics and Biomedicine**, Shanghai, 572-577, 2013.

[49] G. Memis, M. Sert, "Multimodal Classification of Obstructive Sleep Apnea Using Feature Level Fusion", **2017 IEEE 11th International Conference on Semantic Computing (ICSC)**, San Diego, CA, 85-88, 2017.