Investigating the Contribution of Distance-based Features to Automatic Sleep Stage Classification

Ali Abdollahi Gharbali[1, 2], Shirin Najdi[1, 2] and José Manuel Fonseca[1, 2]

[1] CTS, Uninova, 2829-516 Caparica, Portugal

[2] Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa Campus da Caparica, Quinta da Torre, 2829-516 Monte de Caparica, Portugal

{a.gharbali, s.najdi}@campus.fct.unl.pt, jmrf@fct.unl.pt

Abstract

Objective: In this paper, the contribution of distance-based features to automatic sleep stage classification is investigated. The potency of these features is analyzed individually and in combination with 48 conventionally used features. **Methods**: The distance-based set consists of 32 features extracted by calculating Itakura, Itakura-Saito and COSH distances of autoregressive and spectral coefficients of Electrocorticography (EEG) (C_3-A_2), Left EOG, Chin EMG and ECG signals. All the evaluations are performed on three feature sets: distance-based, conventional and total (combined distance based and conventional). Six ranking methods were used to find the top features with the highest discrimination ability in each set. The ranked feature lists were evaluated using *k*-Nearest Neighbor (*k*NN), Artificial Neural Network (ANN), and Decision-tree-based multi-SVM (DSVM) classifiers for five sleep stages including Wake, REM, N1, N2 and N3. Furthermore, the ability of distance-based and conventional features to discriminate between each pair of sleep stages was evaluated using t-test, a hypothesis testing method. **Results**: Distance-based features occupied 25% of top-ranked features. Simulation results showed that using distance-based features together with conventional features can lead to an enhancement of accuracy. The best classification accuracy (85.5%) was achieved by DSVM classifier and 13 features selected by mRMR-MID and normalized with Min-Max method for total feature set, where two of them were from the distance-based feature set. The t-test results show that distance-based features outperform conventional features in discriminating between N1 and REM stages that is usually a challenge for classification systems. **Conclusion**: Distance-based features have a positive contribution to sleep stage classification, including enhancement of accuracy and better REM-N1 discrimination ability. **Significance**: The main motivation for this work was to evaluate new features to characterize each sleep stage in such a way that extracted features were more powerful than conventional features, to distinguish sleep stages from each other, and to improve classifiers accuracy.

Keywords: Polysomnography, Sleep stage classification, Feature extraction, Itakura, Itakura-Saito, Feature selection, Distance-based features.

1. Introduction

Sleep is fundamental for physical and mental health and occupies a significant part of human life. Therefore, the diagnosis of sleep-related disorders is of great importance for sleep research. Normal human sleep consists of two distinct stages with independent functions known as Non-Rapid Eye Movement (NREM) and Rapid Eye Movement (REM) sleep. In an ideal situation, NREM and REM states alternate regularly, each cycle lasting 90 minutes on average. NREM sleep accounts for 75-80% of sleep duration. According to the American Academy of Sleep Medicine (AASM) [1], NREM is subdivided into three stages: stage N1 or light sleep, stage N2 and stage N3 or Slow Wave Sleep (SWS). On the other hand, REM sleep accounts for 20-25% of sleep duration. The first REM state usually occurs 60-90 minutes after the onset of the NREM and lasts a few minutes [2]. To perform sleep analysis, specific physical and electrical activities of the body and brain are recorded. For this aim, a multiple-parametric test, called polysomnography (PSG) is usually used. During the PSG test, a number of biosignals including Electroencephalogram (EEG), Electro-oculogram (EOG), chin Electromyogram (EMG), leg Electromyogram (EMG), airflow signals, respiratory effort signals, oxygen saturation, body position, and electrocardiogram (ECG) are recorded in overnight sleep. The presence of skillful technicians and physicians is necessary for assuring the quality of the recording and analysis. After the acquisition of the PSG, the data are scored by a technician according to a collection of rules set forth by AASM. According to these criteria, the scoring should be performed on 30 second, sequential epochs starting from the first sample of the data. For each stage, a number of recommended definitions are presented in AASM. These definitions mainly include EEG frequency and waveform, eye blinks and movements and EMG amplitude. EEG is divided into five frequency bands: slow wave activity (0.5-2 Hz), delta waves (0-3.99 Hz), theta waves (4-7.99 Hz), alpha waves (8-13 Hz) and beta waves (greater that 13 Hz). As an example, trains of 2-6 Hz saw-tooth waves with high frequency over the central head region in EEG, rapid eye movements in EOG, and low-chin EMG are typical indications of REM stage [1]. Manual scoring of sleep stages has several challenges and disadvantages. First of all, it is time consuming. It usually takes hours to score the PSG of a whole night of sleep. Second, the results of sleep scoring from two different practitioners are often not consistent. It has been reported that there is considerable inter-scorer variability (~ 20% disagreement) among scorers. Such differences are typically the result of rapid transitions between stages, which create ambiguous stages [3]. Moreover, with the immergence of at-home sleep monitoring systems, there is an urgent need for unsupervised methods that can efficiently score sleep data in a way that the results are medically reliable. Therefore, developing automatic sleep stage classification algorithms has been the focus of many researchers.

The common approach in automatic sleep stage classification, like any other pattern recognition process, includes feature extraction and classification steps. Features are extracted from a subset of raw PSG recordings containing only raw EEG data or EEG data together with other raw PSG signals acquired. Various types of features can be extracted from PSG recordings. Since EEG data in a transform domain reveals more useful information than in the time domain, usually a series of transformations such as Fourier transform [4–6], Short Time Fourier Transform (STFT) [7], Wavelet transform [8–12], Hilbert-Hung transform [13], and Empirical Mode Decomposition (EMD) [14–17] are applied. Because the information related to the sleep stages can be inferred from the EEG rhythms, the coefficients resulted from these transforms are used to extract EEG frequency bands. Moreover, these coefficients are often regarded as different representations of the PSG recordings. Therefore, several statistical and nonlinear features are extracted from them [18]. In [4,5] for instance, the mean of the absolute values, average power, and standard deviation are extracted from Discrete Wavelet Transform (DWT) coefficients of each EEG sub-band.

On the other hand, the popularity of the transform-based features does not imply that temporal or nonlinear ones extracted from PSG recordings are not useful in sleep stage classification. Temporal features have lower computational complexity and simulate the manual scoring process performed by the technician. Statistical parameters [4,14,19,20], autoregressive model parameters [21], Hjorth parameters [22,23], and features based on period analysis like zero-crossing rate [20,24] are among the most common temporal features extracted from PSG recordings, especially EEG.

Since the dynamic behavior of individual neurons is governed by threshold and saturation phenomena, nonlinearity is apparent in the brain’s neural network. Moreover, the brain’s ability to perform advanced cognitive tasks rejects the hypothesis of a completely stochastic brain. In addition to the EEG, other signals acquired from the body have neither a completely stochastic nature nor are stationary. Therefore, nonlinear signal processing techniques have also widely been used for characterizing sleep signals. Entropy estimators [23,25–27], fractal dimension [25,28], and Hurst exponent [29,30] are only some of the nonlinear features used in this area. In addition to the conventional features, looking at recent papers in sleep stage classification to understand the latest trends can be useful. There are few works in the area that present new features for enhancing the quality of the sleep staging systems. In [31], two new statistical features were applied to the single-channel EEG: Maximum-Minimum Distance (MMD) and EnergySis (*Esis*). To extract MMD, each EEG epoch is divided into sub-windows. MMD is defined as the Euclidean distance between the maximum and minimum points of the EEG waveform in the corresponding sub-window. MMD of each epoch is calculated by summing the MMDs of the sub-windows. Regarding the second feature, EnergySis, the basic idea is that the signal has energy and speed. This feature is calculated by multiplying the sum of the squared amplitude and velocity of each epoch.

Feature vector quality is an important factor for developing a reliable classification system. Features used in a specific machine-learning problem can perform reasonably well for other problems as well. Therefore, researchers often evaluate and explore the applicability of various features in different machine-learning areas. Kong et al. in [32] assumed that an EEG signal can be modeled as an autoregressive (AR) process and used Itakura distance to measure the similarity of the EEG signals. Itakura distance has, in fact, been found effective in distinguishing hypoxia and asphyxia. Later in 2004, Estrada et al. [33] used the Itakura distance for measuring the similarity of a baseline EEG epoch to the rest of the EEG in the context of sleep stage classification. In addition to the similarity of EEG signal with itself, in [34,35] it is demonstrated that the Itakura distance between EEG and EOG is also a useful similarity measure for sleep stage classification.

In the classification step, various types of classifiers have been used in the literature. Among them, Artificial Neural Network (ANN) [36–39], statistical classifiers such as Support Vector Machines (SVM) [4,9,40,41], instance-based classifiers like *k*-Nearest Neighbor (*k*NN) [40,42] are among the most popular classifiers. In principle, SVMs are designed for binary classification problems, but for the cases that discrimination among more than two classes is required, like sleep stage classification, a multi-class framework of SVM is developed. Several papers provide evidence for the high performance of SVM [4,9,40]. However, several practical challenges such as improving generalization and reducing computational complexity of these systems are still unsolved.

Considering the outstanding performance of Itakura and Itakura-Saito distances in sleep and speech signal processing and COSH distance in speech signal processing, this work extends our initial study on the distance-based features for sleep stage classification [23,45], where Itakura distance outperformed conventional features in classifying the sleep data from the Physionet database [46]. Since the works presented so far use distance-based features in a restricted manner (only the performance of the limited variations are tested), in the current work we aim to extensively evaluate the performance of distance-based features together with conventional features in automatic sleep stage classification. These distance-based features are extracted by calculating Itakura, Itakura-Saito and COSH distances of autoregressive and spectral coefficients of EEG, EMG, EOG and ECG signals.

The following contributions to improve automatic sleep stage classification are presented:

a) Evaluating the potency of distance-based features for sleep stage classification,
b) Comparing the performance of a distance-based feature set with conventionally used features,
c) Assessing the effect of feature normalization on classification results,
d) Utilizing the Vikor method for finding the optimum number of features considering classification accuracy,
e) Analyzing discrimination ability of top features selected by different feature ranking methods, including conventional and distance-based, using statistical hypothesis testing method.

The rest of the paper is structured as follows: Section 2 explains the database used in this work. Section 3 provides a detailed description of the study framework used in this paper. In section 4, the performance of various parts of the framework is assessed and the results are presented. Section 5 finalizes the paper with the conclusions and direction of the future work.

2. Database

In this paper, we used an open-access comprehensive ISRUC-Sleep dataset [47]. This dataset includes data from healthy subjects as well as subjects with sleep disorders and subjects under the effect of sleep medication. PSG recording was performed using a bio-signal acquisition equipment namely, SomnoStar Pro sleep system, in the sleep medicine center of the hospital of Coimbra University (CHUC) between 2009 and 2013. The PSG signals were recorded over a whole night of sleep (approximately eight hours) according to the recommendations of AASM. the sampling frequency was 200 Hz for all EEG, EOG, chin EMG and ECG signals. After segmenting the data into 30 second epochs, two different experts performed manual sleep scoring using AASM. In this dataset, to improve the quality of the recordings a pre-processing step was already taken by the providers of the database.

a) A notch filter was applied to eliminate the 50 Hz electrical noise from EEG, EOG, chin EMG and ECG.

b) EEG and EOG recordings were filtered using a bandpass Butterworth filter with a lower cut-off frequency of 0.3 Hz and higher cut-off frequency of 35 Hz.

c) EMG channels were filtered using a bandpass Butterworth filter with a lower cut-off frequency of 10 Hz and higher cut-off frequency of 70 Hz.

For our evaluations, we used PSG recordings from healthy subjects. Nine male and one female subjects aged between 30 and 58 participated in the recordings. Each recording contains signals from 19 channels. The data include six EEG channels: F3-A2, C3-A2, O1-A2, F4-A1, C4-A1, and O2-A1 from which we selected the C3-A2 EEG channel. The C3-A2 channel is the commonly used EEG channel in sleep stage classification [9,20,48,49] and is among the recommended channels by AASM. In addition to one EEG signal, we used the signals from right EOG and chin EMG, and ECG channels of all ten subjects.

3. Methodology Description

Figure 1 shows the study framework used in this paper. In the following sub-sections, each part will be described in detail.

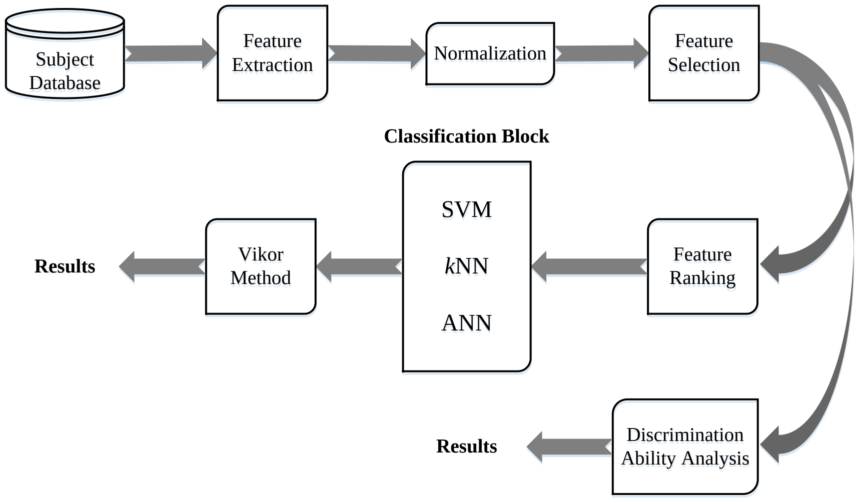


Figure 1. Sleep Study Framework.

3.1. Feature Extraction

In this paper both conventional and new features were extracted from 30-second epochs of PSG data. The details of each group of features are explained as follows.

3.1.1 Conventional Features

The conventional feature vector consists of 48 features extracted from EEG, EOG, and EMG signals. We tried to use the most commonly used features in the literature to explore the information contained in these signals [4,20]. These features can be mainly categorized into temporal, time-frequency domain, entropy-based and non-linear features. To extract time-frequency domain features, a Wavelet Packet (WP) tree with seven levels of decomposition was utilized to extract the EEG rhythms. For more details about EEG rhythms and features refer to [23,50]. As a result, each epoch's feature vector contains 35 EEG, 6 EOG, and 7 EMG features. Table 1 summarizes the conventional features used in this study along with their handy descriptions.

Table 1. Summary of the conventional features extracted from PSG recordings.

Ref.	Signal	Description	T*	TF*	F*	E*	NL*
F1	EEG	Arithmetic Mean	●				
F2		Maximum	●				
F3		Minimum	●				
F4		Standard Deviation	●				
F5		Variation	●				
F6		Skewness	●				
F7		Kurtosis	●				
F8		Median	●				
F9		Petrosian Fractal Dimension					●
F10		Rényi Entropy				●	
F11		Spectral Entropy				●	
F12		Permutation Entropy				●	
F13		Approximation Entropy				●	
F14		Hjorth Parameter (Activity)	●				
F15		Hjorth Parameter (Mobility)	●				
F16		Hjorth Parameter (Complexity)	●				
F17		Mean Curve Length					●
F18		Zero-Crossing Number	●				
F19		Mean Energy					●
F20		Mean Teager Energy					●
F21		Hurst Exponent					●
F22		Mean Quadratic Value of WP Coefficients in Delta Band		●			
F23		Mean Quadratic Value of WP Coefficients in Theta Band		●			
F24		Mean Quadratic Value of WP Coefficients in Alpha Band		●			
F25		Mean Quadratic Value of WP Coefficients in Spindle Band		●			
F26		Mean Quadratic Value of WP Coefficients in Beta1 Band		●			
F27		Mean Quadratic Value of WP Coefficients in Beta2 Band		●			

F28		Mean Quadratic Value of WP Coefficients in All Frequency Bands		●			
F29		F24/(F22+F23)		●			
F30		F22/(F24+F23)		●			
F31		F23/(F22+F24)		●			
F32		F24/F23		●			
F33		F22/F23		●			
F34		Mean of the Absolute Values of WP Coefficients in All Bands		●			
F35		Standard Deviation of WP Coefficients in All Bands		●			
F36	EMG	Spectral Power			●		
F37		Maximum of the Spectral Power Distribution			●		
F38		Mean of the Spectral Power Distribution			●		
F39		Standard Deviation of the Spectral Power Distribution			●		
F40		Temporal Energy					●
F41		Ratio of the Temporal Energy of Current Epoch to The Energy of Previous Epoch					●
F42		Ratio of the Temporal Energy of Current Epoch to the Energy of Next Epoch					●
F43	EOG	Mean	●				
F44		Energy					●
F45		Maximum	●				
F46		Standard Deviation	●				
F47		Skewness	●				
F48		Kurtosis	●				

* T: Temporal, TF: Time-Frequency, F: Frequency, E: Entropy, NL: Non-Linear

3.1.2 Distance-Based Features

The Itakura distance is a very popular distance measure in speech signal processing. Suppose $x(t)$ is the baseline epoch and $y(t)$ is an epoch from the rest of the signal. If we model $x(t)$ and $y(t)$ as AR processes with order *p*, then the vectors $\mathbf{a}_x$ and $\mathbf{a}_y$ would contain the AR coefficients, respectively. Itakura distance of a baseline epoch with others is calculated as:

$$D_I = \ln\left(\frac{\mathbf{a}_y^T \mathbf{R}_x(p)\mathbf{a}_y}{\mathbf{a}_x^T \mathbf{R}_x(p)\mathbf{a}_x}\right). \qquad (1.1)$$

where $\mathbf{R}_x(p)$ and $\mathbf{R}_y(p)$ are the autocorrelation matrixes of $x(t)$ and $y(t)$ with size *p* + 1, respectively. Itakura distance, defined in this way, is asymmetric, i.e. D_I of $x(t)$ and $y(t)$ is not equal to D_I of $y(t)$ and $x(t)$ [51]. To add symmetry to this measure, the mean of these two distances is usually calculated, as [33]:

$$D_I = \frac{1}{2}\left(\ln\left(\frac{\mathbf{a}_y^T \mathbf{R}_x(p)\mathbf{a}_y}{\mathbf{a}_x^T \mathbf{R}_x(p)\mathbf{a}_x} \right) + \ln\left(\frac{\mathbf{a}_x^T \mathbf{R}_y(p)\mathbf{a}_x}{\mathbf{a}_y^T \mathbf{R}_y(p)\mathbf{a}_y} \right) \right). \tag{1.2}$$

In addition to AR coefficients, the distance between spectral representations of the signals can be used to measure similarity [51]. Suppose $S_x(\omega)$ and $S_y(\omega)$ are the power spectra of $x(t)$ and $y(t)$. The Itakura distance between these two spectra, in its asymmetric form, is calculated as:

$$D_I(X,Y) = \ln\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{S_x(\omega)}{S_y(\omega)} d\omega \right]. \tag{1.3}$$

The same averaging (Equation 1.2) can be applied for adding symmetry property to this distance. Along with Itakura distance, there are two other distance measures that are common in speech processing: Itakura-Saito and COSH distances [52]. Following the definitions of variables made for Itakura distance, the Itakura-Saito distance is calculated as:

$$D_{IS}(X,Y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{S_x(\omega)}{S_y(\omega)} - \ln\frac{S_x(\omega)}{S_y(\omega)} - 1 \right] d\omega. \tag{1.4}$$

COSH distance is the symmetrical version of the Itakura-Saito distance and is calculated as:

$$\begin{aligned} D_{Cosh} &= \frac{1}{2}\left(D_{IS}(x,y) + D_{IS}(y,x) \right) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{S_x(\omega)}{S_y(\omega)} + \frac{S_y(\omega)}{S_x(\omega)} - 2 \right) d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} 2\cosh\left(\ln\frac{S_x(\omega)}{S_y(\omega)} - 1 \right) d\omega, \end{aligned} \tag{1.5}$$

where $\cosh(x) = \frac{e^x + e^{-x}}{2}$ is the hyperbolic cosine function. Like the Itakura distance, the Itakura-Saito and COSH distances can be calculated using AR coefficients as well.

Considering the previous work in this area and following the outstanding performance of the Itakura distance in our previous work [23,45], in this study a set of 32 distance-based features were proposed for sleep stage classification as summarized in Table 2. Two types of distance-based features were considered: features measuring the distance of a baseline epoch with other epochs of the same signal and features measuring the distance of a baseline epoch with the epochs of another signal. For calculating F49 to F52 and F73 to F74, the wake EEG epoch was considered as the baseline. The same applies for features F53 to F64 and F75 to F80 corresponding to EMG, EOG, and ECG signals. For calculating F65 to F72, wake EEG epoch was considered as the baseline, and the distance was found between EEG-EOG, EEG-EMG, and EEG-ECG. To the best of our knowledge, except for three features (F49, F51, F65), the remaining features have not been used previously in sleep stage classification. We used a MATLAB speech processing toolbox, namely VOICEBOX [53], consisting of MATLAB routines that are maintained by and mostly written by Mike Brookes, Department of Electrical & Electronic Engineering, Imperial College, UK. We used the routines for calculating Itakura, Itakura-Saito and COSH distances from this toolbox.

Table 2. Summary of distance-based features extracted from PSG recordings

Ref.	Signal	Description
F49	EEG	Itakura Distance of AR Coefficients
F50		Itakura Distance of Spectral Coefficients
F51		Itakura-Saito Distance of AR Coefficients
F52		Itakura-Saito Distance of Spectral Coefficients
F53	EMG	Itakura Distance of AR Coefficients
F54		Itakura Distance of Spectral Coefficients
F55		Itakura-Saito Distance of AR Coefficients
F56		Itakura-Saito Distance of Spectral Coefficients
F57	EOG	Itakura Distance of AR Coefficients
F58		Itakura Distance of Spectral Coefficients
F59		Itakura-Saito Distance of AR Coefficients
F60		Itakura-Saito Distance of Spectral Coefficients
F61	ECG	Itakura Distance of AR Coefficients
F62		Itakura Distance of Spectral Coefficients
F63		Itakura-Saito Distance of AR Coefficients
F64		Itakura-Saito Distance of Spectral Coefficients
F65	EEG & EOG	Itakura Distance of AR Coefficients,
F66		Itakura Distance of Spectral Coefficients
F67		Itakura-Saito Distance of AR Coefficients
F68		Itakura-Saito Distance of Spectral Coefficients
F69	EEG & EMG	Itakura Distance of AR Coefficients
F70		Itakura Distance of Spectral Coefficients
F71		Itakura-Saito Distance of AR Coefficients
F72		Itakura-Saito Distance of Spectral Coefficients
F73	EEG	COSH Distance of AR Coefficients
F74		COSH Distance of Spectral Coefficients
F75	EMG	COSH Distance of AR Coefficients
F76		COSH Distance of Spectral Coefficients
F77	EOG	COSH Distance of AR Coefficients
F78		COSH Distance of Spectral Coefficients
F79	ECG	COSH Distance of AR Coefficients
F80		COSH Distance of Spectral Coefficients

3.2. Feature Normalization

The extracted features from PSG signals are in different ranges, and this variety can bias the results of the following steps. Feature normalization methods are usually utilized for avoiding this bias. In this paper, two different types of normalization methods were used: standardization (or Z-score normalization) and Min-Max. The effect of each method in feature ranking and classification was evaluated. In standardization, the features are rescaled so that they have zero mean and unit variance. In normalization, features are scaled to the fixed range of [0 1]. This rescaling is necessary for many machine learning algorithms.

3.3. Feature Selection

To remove the features with high levels of similarity, a feature selection method was proposed in this paper. Existence of similar features negatively affect the stability [54] of the feature ranking results; therefore, applying feature selection can improve the overall performance of the proposed algorithm [45]. After the L1-norm between each pair of feature vectors was calculated, a similarity threshold was defined. The feature pair, whose L1-norm was less than the threshold level, was considered strongly similar. In this way, the features were clustered into groups of similar features, and one feature per cluster was selected as representative. The representative feature has the lowest computational complexity. Alternatively, it is possible to use Principal Component Analysis (PCA) for finding the most dissimilar features. However, there are two main reasons that we did not use PCA. First, using PCA for finding a non-redundant feature set would lead to keeping and calculating all the features in the classification and practical application steps, whereas by using the similarity threshold, the most redundant features can be detected and omitted from the feature set in the application step. Second, PCA would generate combinations of the features. Since the performance of the distance-based features will be evaluated and compared with the performance of the conventional ones, it is necessary to preserve the information of the features and PCA is not suitable in this regard.

3.4. Feature Ranking

For analyzing the contribution of distance-based features as well as evaluating the potency of the conventional feature set, feature ranking techniques were adopted. In particular, we used ReliefF, minimum Redundancy Maximum Relevance (mRMR-MID and mRMR-MIQ), Fisher score, Chi-square and Information Gain techniques. Next, each technique will be briefly described:

3.4.1 ReliefF

Originally proposed by Kira and Rendell in 1992 [55], Relief [56] is an instance-based method for estimating a features discrimination power. In this method, for a randomly selected sample, the 2*k* nearest neighbors are considered: *k* neighbors from the same class (*hits*) and from a different class (*misses*). Then, the distance of the random sample from the hits and misses is calculated. A quality (discrimination power) coefficient is updated according to this distance, i.e., the feature with lower distance from hits will have higher quality. ReliefF is an extension of the Relief method that removes the two-class problems restriction and reduces its sensitivity to noisy and incomplete data.

3.4.2 minimum Redundancy-Maximum Relevance (mRMR)

This is a feature selection method that selects a subset of features by maximizing the relevance of each feature to the target class and minimizes the redundancy between the selected features. It was mainly proposed by Peng et al. [57] for dealing with the redundancy problem. The redundancy and relevance are calculated using mutual information, whereas the objective function is defined by either the difference between redundancy and relevance (mRMR-MID) or the ratio between relevance and redundancy (mRMR-MIQ).

3.4.3 Fisher Score

This method is one of the most efficient, being widely used for feature ranking. Its main idea is to find a group of features with maximum distance between the data points from different classes and minimum distance between the data points of the same class in the feature space [58]. Since the Fisher score is calculated individually for each feature, the selected feature set can be redundant.

3.4.4 Chi-square Test

This is a statistical test to measure the independency of the events. In feature selection, it is used to evaluate whether or not the occurrence of a specific value of a feature and a specific class are independent. Despite the fact that Chi-square was proposed exclusively for categorical data, this method was later extended to the continuous case [59]. For calculating the Chi-square statistics of each feature, the range of the numerical feature should be discretized into intervals. The features are ranked according to Chi-square statistics without taking into account the interactions between features like Fisher score.

3.4.5 Information Gain (IG)

This method proposed by Ross Quinlan [60] is a widely used feature-ranking algorithm. It works based on a decision tree generated from the training set. To select the effective feature in each node of the tree the IG measure is used. In other words, IG measures how much information each specific feature provides with respect to each class. Therefore, considering the notion of the decision tree, IG depends on how much information was available before knowing the feature and on how much would be available after. A common measure for the information is Shannon entropy, although any measure that allows for evaluating the information content of a feature will be applicable.

Each of these six methods was applied on the conventional, distance-based and total feature sets (combined conventional and distance-based feature sets), and in all, 18 ranked lists of features were achieved.

3.5. Classification

For sleep stage classification three different classifiers were used: *k*NN, ANN and Decision-tree-based multi-SVM (DSVM). The reason for choosing these three different classifiers is that we did not want to restrict the significance of the comparison to one specific family of classifiers, and on the other hand, we aimed to choose a variety of classifiers including the simplest, most used and the one that usually shows the best performance. Euclidean distance was used as the distance measure for the *k*NN classifier. In each experiment, the classification accuracy for the 1, 2, …20 neighborhood was calculated, and the one leading to maximum accuracy was selected as the optimum neighborhood number. For the ANN classifier, a three-layered feed forward neural network with 20 hidden neurons for the conventional and total feature sets and 12 hidden neurons for the distance-based feature set were used.

DSVM was used instead of conventional multi-SVMs. The reason for choosing DSVM was that it outperforms conventional multi-SVMs (OAO and OAA) while utilizing lower number of SVM in the structure [61–64]. Radial Basis Function (RBF) was selected as the kernel function, and sigma was set to 3.0 for the conventional and total feature sets and 1.1 for the distance-based feature set.

For each ranked list of features, created by one of the ranking methods, and each specific classifier, the classification accuracy was calculated for the top 1, 2, … 25 features. Since it is always desirable to achieve the maximum accuracy with the minimum complexity, to find the optimum number of features, a Multi-Criteria Decision Making (MCDM) method called Vikor was used [65].

the Vikor method was originally developed for MCDM problems with contrasting and conflicting criteria. In our case, the accuracy and number of features were two conflicting criteria. This method ranks and selects a set of alternative solutions for the problem at hand, helping decision makers to reach a final decision. The various *J* alternative solutions are denoted as $a_1, a_2, \ldots, a_J$. Suppose that there are *n* criteria, f_{ij} is the value of the i^{th} criterion for j^{th} solution, a_{j}. The compromise ranking is performed by comparing the closeness to the ideal solutions of the criteria (utopian solution F^*). The distance measure of the Vikor method is developed from the L_{p}-metric as:

$$L_{p,j} = \left\{ \sum_{i=1}^{n} \left[w_i \left(f_i^* - f_{ij} \right) / \left(f_i^* - f_i^- \right) \right]^p \right\}^{\frac{1}{p}}, \quad 1 \le p \le \infty; \quad j = 1, 2, ..., J, \tag{1.6}$$

where f_i^* and f_i^- are the best and worst solutions of the i^{th} criterion. After determining the best and worst solutions for all criteria, the Vikor algorithm has the following steps:

1. Compute the values S_j and R_j, j=1, 2, …, J as:

$$S_j = \sum_{i=1}^{n} w_i \left(f_i^* - f_{ij} \right) / \left(f_i^* - f_i^- \right), \tag{1.7}$$

$$R_j = \max_i \left[w_i \left(f_i^* - f_{ij} \right) / \left(f_i^* - f_i^- \right) \right], \tag{1.8}$$

where w_i is the weight of i^{th} criterion expressing its importance.

2. Compute the values Q_j as:

$$\begin{aligned} Q_j &= \upsilon \left(S_j - S^* \right) / \left(S^- - S^* \right) \\ &\quad + (1-\upsilon)\left(R_j - R^* \right) / \left(R^- - R^* \right) \\ &\quad \text{where} \\ S^* &= \min_j S_j, \quad S^- = \max_j S_j, \\ R^* &= \min_j R_j, \quad R^- = \max_j R_j, \end{aligned} \tag{1.9}$$

where υ is the maximum group utility, here $\upsilon = 0.5$.

3. Sort the values of S, R and Q in decreasing order, obtaining three ranked lists.
4. The alternative that minimizes Q is selected as the compromise solution if two conditions of “acceptable advantage” and “acceptable stability in decision making” are satisfied. For more information about these conditions, refer to [66].

3.6. Discrimination Ability Analysis

The ability of the top 25 features in the total feature set, selected by different feature ranking methods, to discriminate between each specific pair of sleep stages was evaluated using two-tailed student’s t-test. These pairs include Wake-REM, Wake-N1, Wake-N2, Wake-N3, REM-N1, REM-N2, REM-N3, N1-N2, N1-N3, and N2-N3. Independent student’s t-test is a hypothesis testing method for comparing the means of two populations.

4. Performance Assessment

In this section, the contribution of distance-based features to sleep stage classification is evaluated when combined with different normalization methods, feature selectors and classifiers using the data described in section 2.

4.1. Evaluation of Feature Selection

The similarity of features was evaluated using the method described in section 3.3 for both conventional and distance-based feature sets. The threshold value for L1-norm between each pair of feature vectors was empirically set to 1e^{-15}.

For conventional and distance-based feature sets, the similar groups were detected and are listed in Table 3.

Table 3. Similar feature groups from the conventional and distance-based feature sets.

Conventional Feature Set	**Group 1**			**Group 2**		
	F36, F38 and F40			F6 and F14		
Distance-based Feature Set	**Group 1**	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**
	F52 and F74	F55 and F75	F56 and F76	F60 and F78	F63 and F79	F64 and F80

According to Table 3, several similar cases were found using this measure. For example, the Hjorth activity parameter is the same as the variation. Furthermore, the COSH distance is the symmetric version of the Itakura-Saito distance. From each group of similar features, one feature with the lowest computational complexity was selected as representative of the group. Therefore, F14, F38 and F40 were removed from the conventional feature set. F74, F75, F76, F78, F79 and F80 were also removed from the distance-based feature set. After removing the redundant features, 45 features remained in the conventional feature set, and 26 features remained in the distance-based feature set.

To assess the usefulness of pruning feature sets, the sleep stage classification accuracy before and after feature selection was evaluated using the conventional, distance-based, and total feature sets. The results obtained by the *k*NN classifier with Euclidean distance are shown in Table 4. The optimum number of neighbors for each case was found (shown in brackets in Table 4) by evaluating the performance of the classifier for different numbers of neighbors. According to the results, removing similar features led to an average improvement of 0.61% for all of the cases. The maximum improvement (2.07%) was observed in the pruning of the conventional feature set using the standardization method. Additionally, it is notable that the accuracy of the classification with the Min-Max method is in all cases higher than the one with the standardization method. This emphasizes the importance of selecting a proper feature normalization method before classification.

Table 4. Classification accuracy for the original and pruned feature sets using the *k*NN classifier. The numbers in brackets refer to the nearest neighbors used in each case. The total feature set is composed of pruned distance-based and pruned conventional feature sets.

Features Normalization	Distance-Based	Pruned Distance-Based	Conventional	Pruned Conventional	Total
STD	60.88 (15)	61.03 (5)	70.90 (15)	72.97 (26)	73.26 (12)
Min-Max	62.30 (10)	62.37 (5)	73.94 (8)	74.10 (8)	74.42 (6)

4.2. Feature Ranking Evaluation

For determining the features that should be given a high priority when dealing with the description of PSG signals, six feature ranking techniques were applied on three feature sets: conventional, distance-based and total feature sets. Furthermore, each feature set was considered with two different normalization methods. From each group, the top 25 features were selected for comparison as shown in Tables 5-7. Table 5 shows the feature ranking results for the conventional feature set. The results of this table are summarized in Figure 2. According to Figure 2, generally temporal and time-frequency domain features are preferred by the ranking methods, whereas frequency domain features are the least preferred ones. Nonlinear and entropy features are always among the top 25 and occupy five to six places on the list. Detailed assessment of the top 25 features leads to the following observations about conventional features:

a) EEG zero-crossing number (**F18**) has been chosen as the best feature by most of the ranking methods with either the standardization or Min-Max method. Even the methods that did not select F18 as the first feature such as ReliefF, have it ranked in the top five best features.

b) Petrosian fractal dimension (**F9**), Hjorth parameter (Mobility) (**F15**), and Hurst exponent (**F21**) are among the top ranked-features by all the methods.

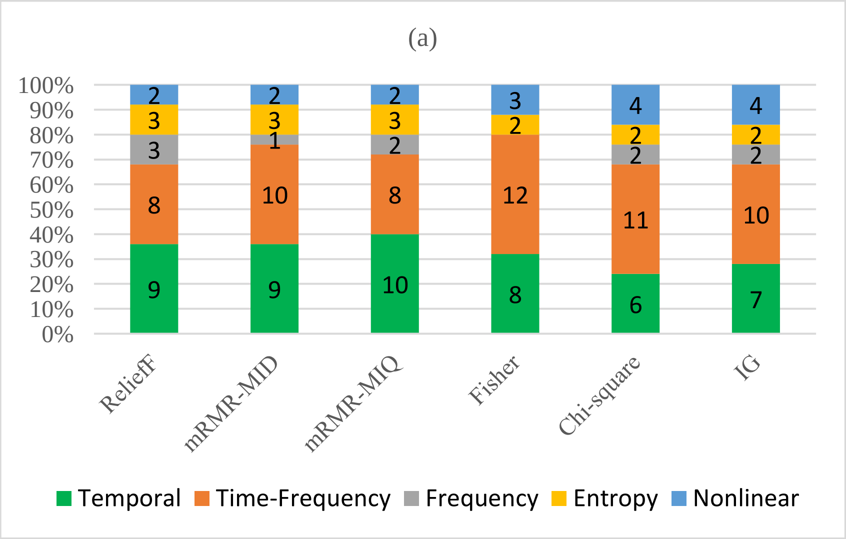
c) ReliefF, mRMR-MID and mRMR-MIQ methods include EEG-, EMG-, and EOG-related features in their top 25 list, whereas Fisher, Chi-square, and IG only contain EEG-related features.

d) Between EMG and EOG features, those related to EOG are more preferred by the ranking methods, such as EOG kurtosis, maximum, and standard deviation.

e) Features from time-frequency domain that were extracted using WP are ranked in the top 25 features by all methods.

Table 5. Feature ranking results for the conventional feature set.

	ReliefF		mRMR-MID		mRMR-MIQ		Fisher		Chi-square		IG	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
1th	F45	F13	F18	F18	F18	F18	F21	F18	F18	F18	F18	F15
2nd	F16	F9	F34	F11	F34	F11	F18	F15	F21	F15	F21	F18
3rd	F15	F21	F10	F9	F10	F45	F15	F21	F15	F21	F15	F21
4th	F18	F15	F21	F35	F37	F35	F9	F9	F9	F9	F34	F9
5th	F13	F18	F35	F21	F29	F9	F13	F13	F34	F16	F9	F16
6th	F29	F16	F15	F45	F13	F32	F34	F34	F35	F11	F35	F11
7th	F21	F32	F13	F15	F23	F31	F35	F35	F4	F26	F4	F2
8th	F9	F29	F29	F32	F21	F10	F11	F16	F28	F13	F28	F13
9th	F32	F45	F23	F31	F45	F21	F4	F4	F22	F2	F23	F34
10th	F7	F7	F46	F10	F35	F30	F16	F25	F16	F27	F22	F22
11th	F31	F31	F9	F13	F15	F15	F29	F29	F23	F20	F5	F35
12th	F48	F6	F26	F30	F25	F29	F22	F30	F36	F22	F19	F3
13th	F41	F25	F11	F29	F11	F34	F30	F22	F5	F34	F11	F26
14th	F6	F10	F4	F34	F48	F13	F28	F33	F19	F25	F16	F4
15th	F25	F41	F25	F4	F9	F23	F25	F28	F11	F29	F36	F20
16th	F11	F48	F2	F25	F2	F25	F33	F27	F27	F3	F13	F27
17th	F10	F46	F31	F23	F26	F4	F31	F26	F13	F30	F27	F29
18th	F36	F11	F16	F33	F32	F33	F23	F31	F29	F35	F29	F30
19th	F39	F42	F32	F16	F31	F16	F2	F2	F20	F4	F2	F36
20th	F46	F34	F37	F2	F4	F2	F5	F5	F26	F36	F30	F28
21th	F27	F3	F45	F22	F46	F22	F27	F19	F30	F33	F20	F25
22nd	F26	F43	F3	F36	F16	F3	F19	F20	F25	F37	F26	F37
23th	F37	F47	F30	F46	F8	F36	F3	F3	F39	F28	F3	F33
24th	F24	F27	F48	F7	F39	F28	F26	F10	F2	F39	F25	F5
25th	F47	F2	F24	F28	F3	F46	F45	F45	F33	F45	39	F19



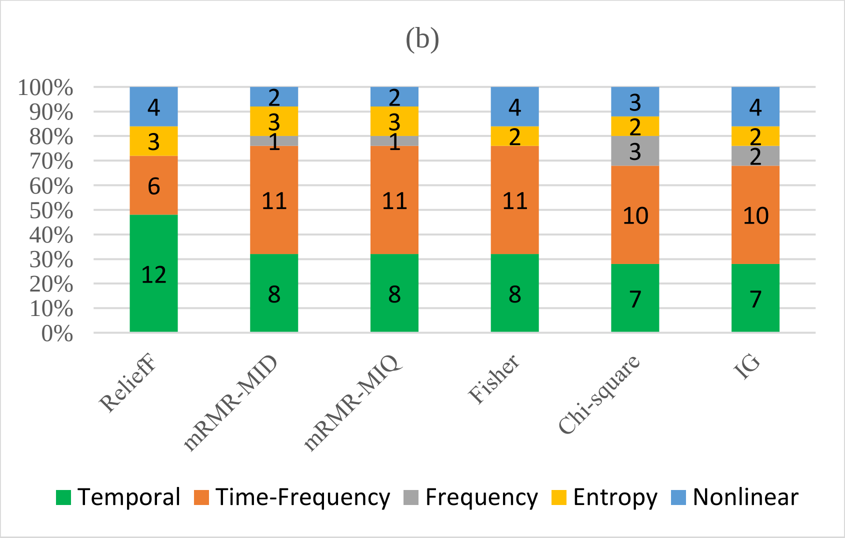


Figure 2. Graphical representation of the feature ranking results for the conventional feature set, (a) normalized with STD and (b) normalized with Min-Max.

Table 6 shows the feature ranking results for the distance-based feature set. Like the conventional feature set, the ranking results are summarized as a graphical representation in Figure 3. According to these charts, Itakura and Itakura-Saito distances were much more effective than COSH distance in discriminating the sleep stages and, at the same time, were preferred equally by the ranking methods. These results imply that the Itakura and Itakura-Saito features can be used interchangeably in sleep stage classification. Detailed assessment of top 25 distance-based features leads to the following observations:

a) Among several types of distance-based features, two are ranked as the best by all methods. These features are similarity between a baseline EEG epoch and the rest of the EEG measured by Itakura distance (**F49** and **F50**) and similarity of EEG and EOG signals measured by either Itakura or Itakura-Saito distance (**F65-F68**).

b) Itakura-Saito distance of AR or spectral coefficients of EEG (**F51** and **F52**) are also seen in the top five.

c) All methods rank one of the features related to the similarity of a baseline EOG epoch to the rest of the EOG (**F57-F60**), measured by Itakura or Itakura-Saito distance, in the top 25.

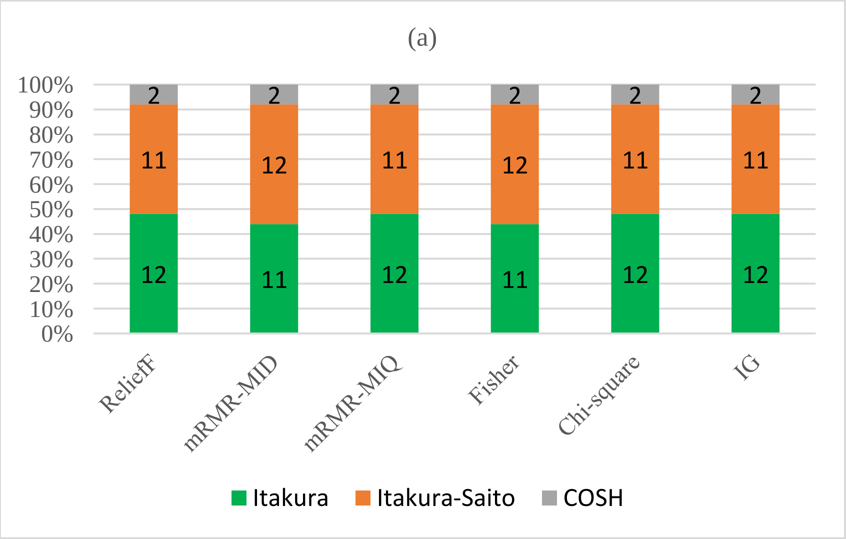
d) The features related to the similarity of a baseline ECG epoch to the rest of the ECG (**F61-F64**), measured by Itakura or Itakura-Saito distance, are considered important mostly by three methods: ReliefF, mRMR-MID and mRMR-MIQ. The same applies to the similarity between EEG and EMG (**F69- F72**).

e) Among the COSH distance-based features (**F73- F80**), only COSH distance of EEG AR coefficients (**F73**) and COSH distance of EOG spectral coefficients (**F77**) could find their way to the top 25 features list.

f) There are no noticeable differences in the number of occurrences of AR or spectral-based features.

Table 6. Feature ranking results for the distance-based feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
1th	F65	F65	F49	F66	F49	F66	F67	F65	F50	F68	F50	F67
2nd	F66	F66	F53	F53	F55	F53	F68	F66	F49	F67	F49	F68
3rd	F49	F50	F68	F67	F68	F67	F66	F68	F51	F49	F51	F49
4th	F50	F49	F61	F57	F58	F57	F65	F67	F73	F50	F73	F50
5th	F70	F70	F58	F61	F65	F49	F54	F53	F52	F51	F52	F51
6th	F69	F69	F65	F49	F61	F55	F53	F54	F57	F73	F58	F73
7th	F61	F58	F69	F69	F53	F65	F56	F55	F58	F52	F57	F52
8th	F62	F57	F55	F68	F50	F68	F55	F56	F67	F57	F67	F57
9th	F72	F72	F50	F65	F67	F69	F49	F49	F68	F58	F68	F58
10th	F71	F71	F67	F55	F57	F54	F50	F50	F65	F65	F65	F65
11th	F52	F62	F71	F51	F66	F61	F57	F57	F66	F66	F66	F66
12th	F73	F61	F57	F63	F54	F51	F58	F58	F60	F60	F59	F60
13th	F51	F60	F59	F54	F69	F50	F70	F70	F59	F59	F77	F59
14th	F63	F77	F66	F59	F51	F70	F69	F69	F77	F77	F60	F77
15th	F64	F59	F54	F52	F56	F56	F51	F73	F53	F53	F53	F53
16th	F57	F63	F70	F71	F63	F52	F73	F51	F54	F54	F54	F54
17th	F58	F52	F51	F64	F59	F58	F52	F52	F55	F55	F55	F55
18th	F60	F51	F63	F56	F73	F63	F60	F60	F56	F56	F56	F56
19th	F77	F73	F72	F50	F70	F73	F77	F77	F61	F70	F61	F70
20th	F59	F64	F56	F70	F52	F59	F59	F59	F62	F69	F62	F69
21th	F55	F53	F60	F73	F60	F64	F72	F72	F63	F72	F63	F72
22nd	F56	F54	F73	F72	F64	F62	F71	F71	F64	F71	F70	F71
23th	F53	F56	F77	F62	F77	F60	F63	F62	F69	F63	F69	F63
24th	F54	F55	F52	F60	F71	F77	F64	F61	F70	F64	F64	F64
25th	F68	F68	F64	F77	F62	F71	F61	F64	F71	F61	F71	F61



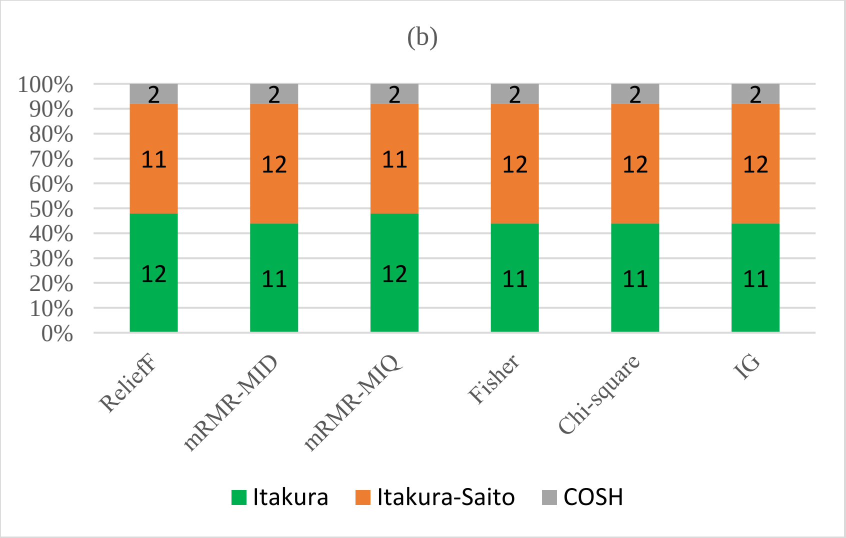


Figure 3. Graphical representation of feature-ranking results for the distance-based feature set (a) normalized with STD and (b) normalized with Min-Max.

Table 7 shows the feature-ranking results for the total feature set. Furthermore, Figure 4 shows the percentage that each feature group occupies in top 25 feature list. Like the conventional feature set, temporal and time-frequency domain features are the most preferred types by the ranking methods. Distance-based features are always in the top 25. Itakura and Itakura-Saito features were more popular than the COSH features. Among the ranking methods, only IG and Chi-square have COSH features in their top 25 feature list.

Detailed assessment of ranking results leads to the following observations:

a) On average, 28% of the top-ranked features was selected from the distance-based feature set. The selected distance-based features in Table 6 belong to one of these categories: similarity of EEG and EOG (**F65-F67**), similarity of a baseline EEG epoch with the rest of EEG (**F49-F52** and **F73**), similarity of a baseline epoch of EMG with the rest of EMG (**F53-F55**), and similarity of a baseline EOG epoch with the rest of EOG (**F57** and **F58**).

b) Among the feature ranking methods, the Chi-square and IG methods had the maximum percentage of distance-based features (44%) in their top 25. These features include the similarity between a baseline EEG epoch with the rest of EEG, measured by Itakura, Itakura-Saito and COSH distances, (**F49-F52** and **F73**) and the similarity of EEG and EOG, measured by the Itakura-Saito distance (**F67** and **F68**).

c) The ReliefF method has the minimum percentage of distance-based features (13%) in its top 25-list. The similarity between EEG and EOG, measured by Itakura distance (**F65** and **F66**), is the selected distance-based feature by this method.

d) **F73** is the only COSH distance-based feature that appears in top 25 list of the total feature set, and it is related to the similarity of a baseline EEG epoch with the rest of EEG.

e) Zero-crossing number (**F18**) is selected as the best feature by all methods.

f) Besides the zero-crossing number, Hjorth parameter (mobility) (**F15**), approximation entropy (**F13**), Petrosian fractal dimension (**F9**), Hurst exponent (**F21**) and at least one of the WP-based features (**F22-F35**) are in the top-ranked features by all methods.

g) There are some features never ranked in the top 25 by any of the methods. Examples of these features are mean curve length (**F17**) and mean Teager energy (**F20**).

Table 7. Feature ranking results for the total feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
1th	F16	F13	F18	F18	F18	F18	F18	F18	F18	F18	F18	F18
2nd	F15	F9	F34	F11	F34	F11	F21	F15	F21	F15	F21	F15
3rd	F13	F21	F53	F9	F53	F45	F15	F21	F15	F21	F15	F21
4th	F29	F29	F21	F66	F48	F53	F9	F9	F9	F9	F34	F9
5th	F32	F32	F35	F35	F68	F9	F13	F13	F49	F16	F9	F16
6th	F45	F16	F68	F21	F32	F35	F34	F65	F50	F68	F35	F49
7th	F18	F15	F15	F45	F46	F32	F67	F66	F34	F67	F4	F50
8th	F7	F7	F46	F15	F21	F66	F68	F34	F35	F49	F49	F68
9th	F9	F18	F13	F31	F35	F31	F66	F35	F4	F50	F50	F67
10th	F21	F31	F29	F32	F13	F21	F65	F68	F51	F11	F22	F11
11th	F65	F11	F23	F53	F10	F10	F35	F67	F73	F51	F23	F51
12th	F66	F6	F2	F13	F11	F30	F11	F16	F52	F73	F28	F73
13th	F10	F45	F57	F10	F25	F15	F54	F4	F22	F52	F52	F52
14th	F6	F10	F11	F4	F23	F23	F53	F54	F16	F13	F51	F2
15th	F48	F34	F9	F29	F15	F29	F4	F53	F28	F26	F73	F13
16th	F41	F25	F26	F30	F58	F13	F16	F25	F23	F2	F5	F34
17th	F36	F47	F4	F65	F55	F34	F29	F30	F11	F27	F19	F22
18th	F31	F66	F55	F34	F2	F67	F30	F29	F68	F20	F11	F3
19th	F39	F65	F65	F23	F29	F25	F25	F33	F67	F22	F68	F35
20th	F37	F24	F49	F25	F9	F4	F33	F22	F36	F34	F67	F26
21th	F61	F48	F25	F54	F26	F65	F31	F28	F58	F65	F16	F4
22nd	F62	F41	F31	F33	F65	F33	F56	F27	F5	F66	F58	F66
23th	F2	F37	F10	F67	F4	F54	F55	F31	F57	F3	F57	F65
24th	F34	F46	F67	F68	F37	F68	F22	F5	F19	F25	F13	F57
25th	F46	F43	F32	F49	F31	F69	F27	F26	F13	F29	F36	F58

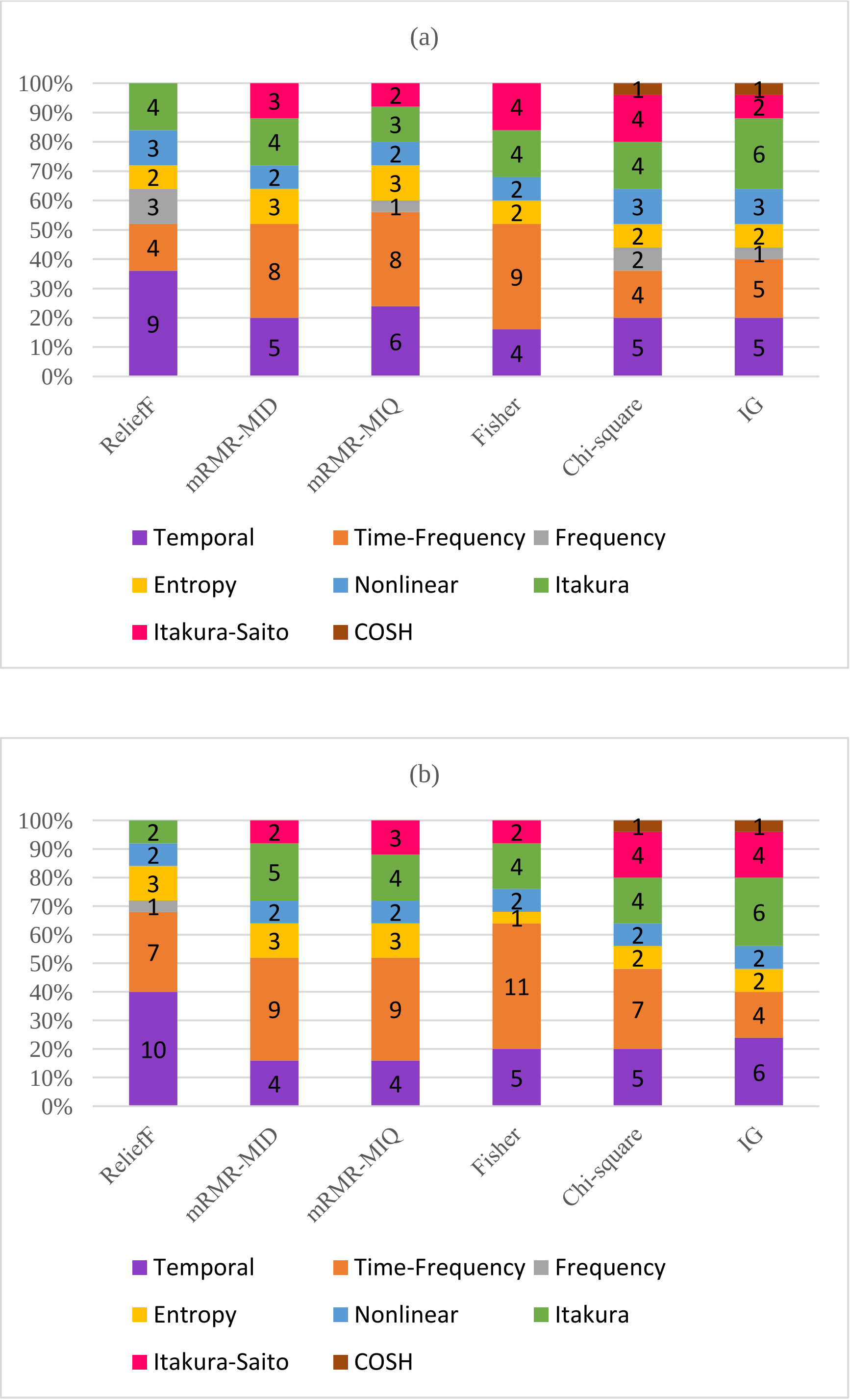


Figure 4. Graphical representation of feature ranking results for the total feature set (a) normalized with STD (b) normalized with Min-Max.

4.3. Evaluation of the Classification Results

Tables 8-16 demonstrate the 5-stage (Wake, REM, N1, N2 and N3) classification accuracy results along with the optimum number of features selected by the Vikor method for all three feature sets and three classifiers. The reliability of the results was validated by using 10 times repeated 10-fold cross validation method on the whole data from 10 healthy subjects. Simulations were performed using a PC with 3.40 GHz Intel® Core™ i7-3770 CPU, 8 GB of RAM, Windows 10 (64 bits), and MATLAB R2015b. For each ranked list of features, created by one of the ranking methods, and each classifier, the overall classification accuracy, sensitivity and specificity were calculated for the top 25 features. Analyzing the results reveals that, starting with one feature, each additional feature typically leads to an increment in the classification accuracy.

However, at some point, the increment on the classification accuracy for each additional feature is not significant. Inspired by MCDM problems, in this paper the Vikor method was applied to the classification results for determining the optimal feature number that provides a satisfactory trade-off between the selected number of features and the classification accuracy. This method is one of the most common MCDM techniques with straightforward calculations. Accuracy and number of features were two conflicting criteria with the corresponding weights of 0.7 (w_1) and 0.3 (w_2), respectively, meaning that, in our sleep stage classification system, classification accuracy had priority over complexity. Figure 5 shows a sample of the Vikor method results for the features scaled by standardization method, ranked with ReliefF and classified by *k*NN classifier. The utopian solution, shown with a black star, represents the ideal solution in which the accuracy is maximum and the number of features is minimum. The selected point by the Vikor method in each case is the closest point of the Pareto front (the set of solutions) to the utopian solution considering the weights of the two criteria.

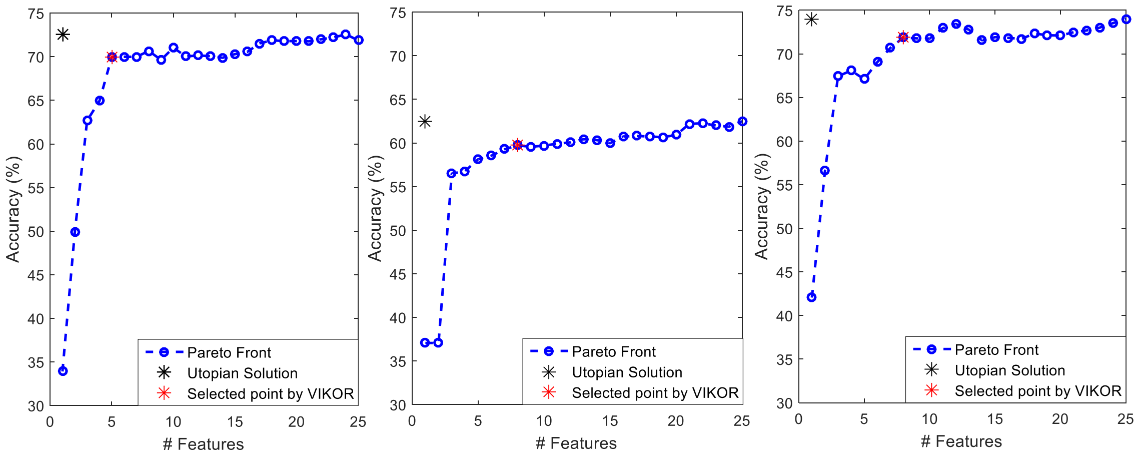


Figure 5. Optimum number of features selected by the VIKOR method for the (a) conventional, (b) distance-based, and (c) total feature sets.

4.3.1 *k*NN Classifier Results

Next, assessment of the results related to the *k*NN classifier (Tables 8-10) will be discussed.

a) The maximum enhancement in classification accuracy after adding the distance-based features to the conventional feature set occurred in mRMR-MID with Min-Max.

b) For all three feature sets, the maximum accuracy, regardless of the feature normalization method, was achieved by mRMR-MID or mRMR-MIQ method. Seven and in one case eight features were selected by the Vikor method to achieve this accuracy. The Itakura distance of EEG-EOG spectral coefficients, Itakura-Saito distance of EEG-EOG spectral coefficients, and Itakura distance of EMG AR coefficients are among these features.

c) For all three feature sets, the minimum accuracy, regardless of the feature normalization method, was achieved by the Chi-square method.

d) For most of the ranking methods, adding distance-based features to the conventional feature set improved the sensitivity and specificity of the classification.

Table 8. *k*NN classifier results for the conventional feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	5	6	10	7	9	8	8	8	7	8	8	8
# Neighbors	18	16	20	11	20	20	12	12	12	20	16	8
Sensitivity	72.8	72.9	75.6	72.5	73.7	71	71.5	72.7	71.3	74.6	73.1	72.9
Specificity	93.4	93.2	94	93.4	93.5	92.6	93.1	93.2	92.9	93.8	93.3	93.4
Accuracy	70	70.9	72.1	71.3	72.9	70.8	69.7	71.6	69	71.9	69.2	72.7

Table 9. *k*NN classifier results for the distance-based feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	8	11	6	6	6	5	11	12	10	8	10	8
# Neighbors	19	6	16	9	17	9	10	11	18	12	17	12
Sensitivity	64.3	61.7	62.5	65.6	64	63	64.3	63.3	63.9	60	64.7	61.5
Specificity	91.2	90.6	90.4	91.1	90.6	91	91.5	90.8	91	89.9	91	90.3
Accuracy	59.7	59	61.5	60.6	61.9	60	62	60	61	56.3	61.1	56.6

Table 10. *k*NN classifier results for the total feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	8	8	8	7	7	7	7	6	7	8	8	10
# Neighbors	14	10	11	6	17	12	10	11	11	10	19	10
Sensitivity	75.1	73.3	74.1	77.4	75.1	75.3	76.5	73.4	72.3	70.6	74	75.4

Specificity	93.8	93.7	93.6	94.2	93.9	93.6	94.2	93.5	93	92.4	93.8	94.3
Accuracy	72	71	73.2	73	72.2	72.3	71.1	71	71	70	71	70.3

4.3.2 ANN Classifier Results

Next, assessment of results related to the ANN classifier (Tables 11-13) will be discussed.

a) The maximum enhancement in classification accuracy after adding the distance-based features to the conventional feature set occurred in mRMR-MIQ with standardization.

b) For all three feature sets, the maximum accuracy, regardless of feature normalization method, was achieved by the mRMR-MID or mRMR-MIQ method. Up to 11 features were selected by the Vikor method to achieve this accuracy. The Itakura distance of the EEG-EOG spectral coefficients, Itakura-Saito distance of the EEG-EOG spectral coefficients, and Itakura distance of the EMG AR coefficients are among these features.

c) Compared to the results of the *k*NN classifier, the overall accuracy, sensitivity and specificity of the ANN classifier is higher for three feature sets.

Table 11. ANN classifier results for the conventional feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	9	10	11	11	9	11	8	10	9	9	11	8
Sensitivity	72.6	77.7	75.9	78.3	74.9	76	73.9	74.6	73.4	76.9	73.6	75.4
Specificity	93.7	94.4	94	94.6	93.7	94	93.5	93.6	93.3	94.2	93.4	93.9
Accuracy	79	80	80	80.6	79	79.8	79.8	79.2	78.5	79.7	78.7	79.6

Table 12. ANN classifier results for the distance-based feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	9	7	7	7	13	7	11	13	15	15	15	15
Sensitivity	62.1	59.9	63.3	61.1	64.8	61.3	63.4	63.6	66.1	64	65.1	63
Specificity	90.5	90	90.9	90.2	91.1	90.3	90.8	90.9	91.5	90.5	91.2	90.7
Accuracy	74.3	72.1	75.2	74	75.6	74	75	74.2	75	73.1	75	73.1

Table 13. ANN classifier results for the total feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	8	9	9	10	8	11	9	9	9	14	8	10
Sensitivity	75.1	75.4	76.5	76.7	76.7	78.8	74	74.8	73.3	76.3	74	74.2
Specificity	93.8	93.8	94.1	94.3	94.2	94.7	93.5	93.7	93.3	94.1	93.5	93.5
Accuracy	79.5	79.2	80.2	79.9	80.2	80.4	79.2	79.1	79.2	79.5	79.2	78.5

4.3.3 DSVM Classifier Results

Next, assessment of results related to the DSVM classifier (Tables 14-16) will be discussed.

a) The maximum enhancement in classification accuracy after adding the distance-based features to the conventional feature set occurred in mRMR-MIQ with Min-Max.

b) For all three feature sets, the maximum accuracy, regardless of the feature normalization method, was achieved by the mRMR-MID or mRMR-MIQ methods. Up to 13 features were selected by the Vikor method to achieve this accuracy. The Itakura distance of the EEG-EOG spectral coefficients, Itakura-Saito distance of the EEG-EOG spectral coefficients, and Itakura distance of the EMG AR coefficients are among these features.

c) Considering that the overall performance of the DSVM classifier, including accuracy, sensitivity and specificity, is the highest among the classifiers used in this paper, it can be concluded that DSVM outperforms *k*NN and ANN classifiers in sleep stage classification.

Looking at the results for all the classifiers, the accuracy obtained by Min-Max is higher than standardization in most of the cases. Furthermore, the presence of the distance-based features among selected features by the Vikor method shows their positive contribution to sleep stage classification.

Table 14. DSVM classifier results for the conventional feature set.

	ReliefF		mRMR-MID		mRMR-MIQ		Fisher		Chi-square		IG	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	10	10	10	11	9	9	8	11	8	9	8	8
Sensitivity	79.2	74.4	80.1	78.5	79	76.3	77.2	76.6	73.2	78.4	76.3	75.7
Specificity	95.3	94.2	95.7	94.9	95.6	94.6	95.2	94.6	94.7	95.4	94.9	94.7
Accuracy	83.7	84.5	84.0	84.7	84.0	83.8	81.5	81.7	81.0	81.9	81.0	81.8

Table 15. DSVM classifier results for the distance-based feature set.

	ReliefF		mRMR-MID		mRMR-MIQ		Fisher		Chi-square		IG	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	9	11	7	6	8	6	11	11	9	15	15	15
Sensitivity	61.1	60.6	70.1	63.6	70.3	60.7	64.1	58.3	62.3	62.9	68.5	64.4
Specificity	91.1	90.9	93.4	92.1	93.4	91.1	91.8	90.7	91.7	91.5	92.8	92.5
Accuracy	78.1	77.2	79.7	79.3	79.8	77.8	79.2	78.1	77.8	78.7	79.4	79.2

Table 16. DSVM classifier results for the total feature set.

	ReliefF		mRMR-MID		mRMR-MIQ		Fisher		Chi-square		IG	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	11	9	8	13	8	11	9	14	9	14	9	15
Sensitivity	79.3	76	81.6	79.8	80.6	80.5	75.1	76.3	75.3	73.8	77.5	76.5
Specificity	95.5	94.9	96.5	96.3	96.1	96	94.6	95.3	94.6	94.3	94.9	94.8

Accuracy	84.8	82.0	84.4	85.5	84.7	85.3	81.3	81.9	80.8	81.6	80.8	81.7

4.4. Evaluation of Discrimination Ability Analysis

As mentioned in section 3.6, to perform a comprehensive analysis and compare the discrimination ability of conventional and distance-based feature sets, independent t-tests were applied on the top 25 features of the total feature set (according to Table 7) with standardization and Min-Max methods. The significance level (α-value) for the t-test was chosen as 0.05, which is a common value. Tables 17 and 18 present the results. In these tables, two categories of features are noticeable, namely “Discriminative” and “Redundant”. These categories are defined as:

a) **Discriminative**: Features with the highest discrimination ability between corresponding pairs of stages were included in this category. From the perspective of the t-test results, features with the lowest *p*-value were categorized as “Discriminative” features.

b) **Redundant**: Features that cannot discriminate between corresponding pairs of stages were included in this category. From the perspective of the t-test results, features with a *p*-value of more than 0.05 were categorized as “Redundant” features.

Table 17. Discrimination ability analysis results for the top 25 features selected from the total feature set with standardization.

	“Discriminative” Features	**“Redundant” Features**
Wake-REM	**F13**, F15, F18, F21, F53, F54, F55, F56.	F6, F31, F41, F61, F62, F67, F68.
Wake-N1	**F13**, F15, F18, F21, F25, F34, F45, F46.	F6, F29, F41, F49, F50.
Wake-N2	F9, F13, **F15**, F18, **F21**.	F6, F23, F30, F33.
Wake-N3	F9, F13, F15, **F18**, F65, F66.	F2, F6.
REM-N1	F13, F15, F18, F21, **F53**, **F54**, F55, F56.	F5, F6, F19, F22, F41.
REM-N2	F2, F4, F23, F26, **F34**, F35, F53, F54, F55, F56, F65, F66.	F6, F41, F51, F52, F73.
REM-N3	F2, **F4**, F5, F9, F11, F15, F18, F19, F21, F22, F23, F28, F29, F31, F36, F65, F66.	F10, F27, F36, F41, F46, F61, F62.
N1-N2	F4, F9, **F11**, F15, F18, F23, F29, F34, F35.	F6, F36, F45, F46, F55, F56.
N1-N3	F4, F5, F9, **F11**, F15, F16, F18, F19, F21, F22, F23, F28, F29, F30, F31, F33, F34, F35, F49, F50, F65, F66.	F26, F36, F39, F41.
N2-N3	F4, F5, F9, F11, F15, **F18**, F21, F29, F30, F31, F33, F34, F35, F46.	F25, F36, F37, F39, F41, F61, F62.

Table 18. Discrimination ability analysis results for the top 25 features selected from the total feature set with min-max.

	"Discriminative" Features	"Redundant" Features
Wake-REM	F9, **F13**, F15, F18, F21, F53, F54.	F6, F31, F43, F47, F48, F67, F68.
Wake-N1	F13, F15, F18, F21, **F25**, F34, F45.	F6, F29, F41, F43, F47, F49, F50.
Wake-N2	F9, **F15**, F18, F21.	F7, F10, F30, F33, F43, F47.
Wake-N3	F9, **F15**, F18, F21, F29, F65, F66.	F2, F3, F6, F24, F43, F47.
REM-N1	F13, F15, F21, **F52**, **F53.**	F6, F22, F41, F43, F47, F51, F52, F71, F72, F73.
REM-N2	F2, F3, F10, F34, F35, F53, F54, **F65**, **F66**.	F11, F21, F32, F41, F43, F47.
REM-N3	F2, F3, F4, F9, F15, F18, F21, F34, F35, **F65**, **F66**.	F6, F10, F42, F43, F47.
N1-N2	F9, F13, F21.	F6, F12, F20, F25, F43, F45, F46, F47, F48, F51, F52, F73.
N1-N3	F4, F9, F15, **F18**, F21, F29, F30, F31, F32, F34, F35.	F6, F26, F43, F47.
N2-N3	F15, **F18**, F21.	F25, F41, F43, F47.

The features with highest discrimination ability (minimum *p*-value) are shown in bold. Assessment of the results in Tables 17 and 18 leads to the following observations:

a) The minimum number of "Redundant" group features is related to the Wake-N3 pair with two features in the standardization method.

b) The maximum number of "Redundant" group features is related to the N1-N2 pair with 11 features in the Min-Max method.

c) The maximum number of "Discriminative" group features is related to the N1-N3 pair with 22 features in the standardization method.

d) The minimum number of "Discriminative" group features is related to the N2-N3 pair with three features in the Min-Max method.

e) There were some features in the Min-Max method that could not distinguish between any of the sleep stage pairs and were always categorized in the "Redundant" group, such as F43 and F47.

f) There were some features that could always distinguish between any pair of sleep stages and were always categorized in the "Discriminative" group. For the standardization method, these features were: F4, F7, F9, F11, F13, F15, F18, F21, F28, F32, F34, F35, F44, F53, F54, F57, F58, F65 and F66 (19 features in total). The distance-based features constitute 31% of these features. For the Min-Max method, the features always categorized as "Discriminative" include: F4, F5, F9, F13, F15, F16, F18, F19, F27, F28, F34, F35, F36, F37, F39, F44, F53, F54, F57, F58, F65, F66, F69, and F70 (24 features in total). The distance-based features constitute 33% of these features

g) Among distance-based features, the Itakura distance of EEG-EOG (F65 and F66) has the highest discrimination ability for both of the normalization methods.

4.5. GPU vs. CPU performance analysis for ANN

Considering the computational complexity of ANN, GPU was utilized in this work to accelerate the simulation process, and significant differences compared to the CPU were noted. The MATLAB parallel computing toolbox was used for distributing the process among several sessions in the computer. A maximum of eight local workers can be provided by the parallel computing toolbox on a single workstation [67]. Nevertheless, the number of usable workers for a process is dependent on the number of cores of the processor in the computer. In our case, a core-i7 processor provided four workers considering the version of MATLAB. An NVIDIA GeForce GT 640 card was used as GPU. The elapsed time for 20 neurons of ANN using CPU was 4426.57 seconds and GPU was 1116.51 seconds. Therefore, it can be concluded that, for this type of computation, GPU is approximately 4-times faster than CPU.

5. Discussion and Conclusions

In this work, the main motivation was to evaluate new features’ ability in characterizing sleep stages in a way that they outperform the conventional ones in distinguishing sleep stages from each other and also to improve the classification accuracy. The performance of the distance-based feature set along with 48 conventional temporals, frequency domain, time-frequency domain, non-linear, and entropy-based features were evaluated in sleep stage classification. The distance-based feature set included 32 features that were extracted by calculating the distance between AR and spectral coefficients of EEG, EOG, EMG, and ECG signals. The distance measures used were Itakura, Itakura-Saito and COSH, all common in speech signal processing. Extensive assessments were performed to reveal the weaknesses and strengths of these features to present a vision for future sleep research.

Similar features were removed from the feature sets through thresholding L1-norm between feature vectors. This step was advantageous because removing these features reduces the final feature vector dimensionality and enhances the stability of feature-ranking results. Moreover, according to the results of Table 4, this step led to an improvement in the classification accuracy. This improvement was expected since the existence of redundant features has no positive effects on the classification results and increases the computational complexity of the whole system.

To find the most useful features for describing PSG signals, six feature-ranking methods, namely ReliefF, mRMR-MID, mRMR-MIQ, Chi-square, Fisher, and IG, were applied to the three feature sets (conventional, distance-based and total). According to the ranking results, from the conventional feature set, EEG zero-crossing rate was selected as the best feature by most of the ranking methods. This result is consistent with the results presented in our previous study using the Physionet sleep database [46] as well as the results of [5] where several features were evaluated to find the optimum feature set for online sleep stage classification. In addition to the zero-crossing rate, Petrosian fractal dimension, Hjorth mobility parameter, and Hurst exponent were always among the top-ranked features. This validates the outstanding performance of these features already demonstrated in previous studies such as [20,45].

In [34,35], it had been shown that the Itakura distance between EEG and EOG signals and also between a reference EEG epoch and other EEG epochs has meaningful variations in different sleep stages. It was concluded that these measures can be used as useful features in automated sleep staging systems and our simulations confirmed this conclusion. According to the results, all the ranking methods listed EEG Itakura distance, EEG-EOG Itakura and Itakura-Saito distances in their top 25 features. Moreover, the features related to the similarity of a baseline EOG/EMG epoch to the rest of the EOG/EMG were always among the top 25 features.

The ranking results for total feature set in Table 7 show that the top 25 features for all the ranking methods include features from both conventional and distance-based sets. This fact implies that a combination of features from different domains yields better results. According to this table, distance-based features occupy 28% of the top-ranked features. These features belong to one of the following categories: similarity of EEG and EOG (F65-F67), similarity of a baseline EEG epoch with the rest of EEG (F49-F52 and F73), similarity of a baseline epoch of EMG with the rest of EMG (F53-F55) and similarity of a baseline EOG epoch with the rest of EOG (F57 and F58). From the conventional features, again zero-crossing rate, Hjorth parameter (mobility), approximation entropy, Petrosian fractal dimension, Hurst exponent and at least one of WP-based features are in the top-ranked features by all methods.

To further investigate the contribution of distance-based features, three different classifiers, *k*NN, ANN and DSVM, were used. Previous studies [5,68] showed that combining different types of features, i.e. temporal, spectral, time-frequency domain and nonlinear, would lead to a satisfactory level of classification accuracy with a fewer number of features. In this paper, we showed that using distance-based features together with conventional ones can further improve the performance of the sleep scoring system. This improvement is noticeable in the results of all three classifiers. According to the results of the Vikor method, 8-13 carefully selected measures from the total feature set are sufficient to reach, on average, 85% accuracy, and usually three of these features are from the distance-based category. The only method that listed conventional features higher in rank than distance-based features is the ReliefF method. Specifically, with Min-Max normalization, this method has its first distance-based feature ranked 18th.

Regarding the compatibility of feature ranking and classifier, all classifiers achieved the highest accuracy with either mRMR-MID or mRMR-MIQ. In particular, mRMR-MID with Min-Max normalization gave the highest accuracy with 13 features in which the EEG-EOG Itakura distance of spectral coefficients and EMG Itakura distance of autoregressive coefficients are selected from the distance-based feature set (Table 16). This result is consistent with our previous study in which we used the Physionet sleep database [46] to evaluate the applicability of rank aggregation to the sleep scoring problem. Moreover, simulation results showed that DSVM outperformed the other two classifiers for all the feature ranking and normalization methods.

According to the literature [69], there has been a lack of discriminative features for distinguishing N1 stage from other sleep stages because neurophysiological signals of N1 and N2 have similarities with each other as well as other sleep stages [70]. For example, the PSG recordings show similar wave patterns in REM and N1 in EEG, both having low amplitude waves of 3-7 Hz [1]. Therefore, the accuracy obtained for classifying the N1 stage is usually lower than for the other stages, and discriminating N1 from REM is especially challenging. To tackle this challenge and increase the discrimination ability of the overall system, other channels (EOG, EMG and ECG) along with EEG are usually used [37,68,71]. In this paper, the ability of the features to discriminate between each pair of sleep stages was assessed using two-tailed student’s t-test. This test was applied on the total feature set. The t-test results show that distance-based features outperform conventional features in discriminating between N1 and REM stages. According to Tables 17 and 18, the Itakura-Saito distance of EEG spectral coefficients (F52) and Itakura distance of EMG spectral and AR coefficients (F53 and F54) have outstanding performances in distinguishing N1 from REM stage, regardless of the feature normalization method. Therefore, these features can be appropriate choices to be included in the sleep stage classification feature set to increase the discrimination ability of the system. Regarding the effect of feature normalization on the overall performance, results show that the Min-Max method slightly outperforms standardization. In other words, the accuracy achieved with the data normalized by Min-Max turned out to be higher than the accuracy achieved with standardization. To obtain a more general conclusion, the effect of feature normalization should be evaluated with different sleep databases.

Despite the advantages, there were some limitations in this work. The studied sleep stage classification system was based on a hypnogram created from the consensus of two experts on visual sleep scoring. There were some cases of interscorer variability, especially on N1. Moreover, the database was pre-processed, and raw data were not available for possible changes in the pre-processing step. Finally, this system is designed for classification of data acquired from healthy subjects. Generalization of the results to unhealthy or elderly subjects would require modifications in the classification rules. Future work can include the analysis of other databases as well as different classifiers for further investigation. Furthermore, comparing the performance of handcrafted features with automatically extracted features by deep networks can be useful.

Acknowledgments

This work was partially funded by FCT Strategic Program UID/EEA/00066/203 of UNINOVA, CTS.

References

[1] R.B. Berry, R. Brooks, C.E. Gamaldo, S.M. Harding, R.M. Lioyd, C.L. Marcus, B. V. Vaughn, AASM - Manual for the Scoring of Sleep and Associated Events version 2.1, (2014).

[2] S. Chokroverty, Sleep Disorders Medicine: Basic Science, Technical Considerations, and Clinical Aspects, Saunders/Elsevier, 2009.

[3] H. Danker-Hopfe, P. Anderer, J. Zeitlhofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, S. Parapatics, B. Saletu, A. Schmidt, G. Dorffner, Interrater reliability for sleep scoring according to the Rechtschaffen &amp; Kales and the new AASM standard., J. Sleep Res. 18 (2009) 74–84. doi:10.1111/j.1365-2869.2008.00700.x.

[4] B. Koley, D. Dey, An ensemble system for automatic sleep stage classification using single channel EEG signal, Comput. Biol. Med. 42 (2012) 1186–1195. doi:10.1016/j.compbiomed.2012.09.012.

[5] M. Radha, G. Garcia-Molina, M. Poel, G. Tononi, Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal, in: 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE, 2014: pp. 1876–1880. doi:10.1109/EMBC.2014.6943976.

[6] H.G. Jo, J.Y. Park, C.K. Lee, S.K. An, S.K. Yoo, Genetic fuzzy classifier for sleep stage identification, Comput. Biol. Med. 40 (2010) 629–634. doi:10.1016/j.compbiomed.2010.04.007.

[7] D. Görür, U.H. Halıcı, G. Ongun, F. Özgen, K. Leblebicioğlu, Sleep Spindles Detection Using Autoregressive Modeling, Proc. ICANN/ICONIP. (2003).

[8] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, H. Dickhaus, Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier, Comput. Methods Programs Biomed. 108 (2012) 10–19. doi:10.1016/j.cmpb.2011.11.005.

[9] J. Kim, A Comparative Study on Classification Methods of Sleep Stages by Using EEG, J. Korea Multimed. Soc. 17 (2014) 113–123. doi:10.9717/kmms.2014.17.2.113.

[10] A.R. Hassan, M.I.H. Bhuiyan, A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features, J. Neurosci. Methods. 271 (2016) 107–118. doi:10.1016/j.jneumeth.2016.07.012.

[11] A.R. Hassan, A. Subasi, A decision support system for automated identification of sleep stages from single-channel EEG signals, Knowledge-Based Syst. 128 (2017) 115–124. doi:10.1016/j.knosys.2017.05.005.

[12] A.R. Hassan, M.I.H. Bhuiyan, Dual tree complex wavelet transform for sleep state identification from single channel electroencephalogram, in: 1st IEEE Int. Conf. Telecommun. Photonics, ICTP 2015, 2016. doi:10.1109/ICTP.2015.7427924.

[13] Yi Li, Fan Yingle, Li Gu, Tong Qinye, Sleep stage classification based on EEG Hilbert-Huang transform, in: 2009 4th IEEE Conf. Ind. Electron. Appl., IEEE, 2009: pp. 3676–3681. doi:10.1109/ICIEA.2009.5138842.

[14] A.R. Hassan, M.I.H. Bhuiyan, Computer-aided sleep staging using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and bootstrap aggregating, Biomed. Signal Process. Control. 24 (2016) 1–10. doi:10.1016/j.bspc.2015.09.002.

[15] A.R. Hassan, M.I. Hassan Bhuiyan, Automatic sleep scoring using statistical features in the EMD domain and ensemble methods, Biocybern. Biomed. Eng. 36 (2016) 248–255. doi:10.1016/j.bbe.2015.11.001.

[16] A.R. Hassan, M.I.H. Bhuiyan, Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting, Comput. Methods Programs Biomed. 140 (2017) 201–210. doi:10.1016/j.cmpb.2016.12.015.

[17] A.R. Hassan, M.I.H. Bhuiyan, Automatic sleep stage classification, in: 2015 2nd Int. Conf. Electr. Inf. Commun. Technol., IEEE, 2015: pp. 211–216. doi:10.1109/EICT.2015.7391948.

[18] A.R. Hassan, S. Siuly, Y. Zhang, Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating, Comput. Methods Programs Biomed. 137 (2016) 247–259. doi:10.1016/j.cmpb.2016.09.008.

[19] S.T.-B. Hamida, B. Ahmed, Computer Based Sleep Staging: Challenges for the Future, in: 2013 7th IEEE GCC Conf. Exhib., IEEE, 2013: pp. 280–285. doi:10.1109/IEEEGCC.2013.6705790.

[20] B. Şen, M. Peker, A. Çavuşoğlu, F. V. Çelebi, A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms, J. Med. Syst. 38 (2014) 18. doi:10.1007/s10916-014-0018-0.

[21] R. Kaplan, Y. Wang, K. Loparo, M. Kelly, Evaluation of an automated single-channel sleep staging algorithm, Nat. Sci. Sleep. 7 (2015) 101. doi:10.2147/NSS.S77888.

[22] B. Hjorth, EEG Analysis Based on Time Domain Properties, Electroencephalogr. Clin. Neurophysiol. 29 (1970) 306–310. doi:10.1016/0013-4694(70)90143-4.

[23] S. Najdi, A.A. Gharbali, J.M. Fonseca, A Comparison of Feature Ranking and Rank Aggregation Techniques in Automatic Sleep Stage Classification Based on Polysomnographic Signals, in: 4th Int. Conf. IWBBIO, Springer International Publishing, Granada, 2016: pp. 230–241. doi:10.1007/978-3-319-31744-1_21.

[24] J. Virkkala, J. Hasan, A. Värri, S.-L. Himanen, K. Müller, Automatic sleep stage classification using two-channel electro-oculography, J. Neurosci. Methods. 166 (2007) 109–115. doi:10.1016/j.jneumeth.2007.06.016.

[25] J. Rodríguez-Sotelo, A. Osorio-Forero, A. Jiménez-Rodríguez, D. Cuesta-Frau, E. Cirugeda-Roldán, D. Peluffo, Automatic Sleep Stages Classification Using EEG Entropy Features and

Unsupervised Pattern Analysis Techniques, Entropy. 16 (2014) 6573–6589. doi:10.3390/e16126573.

[26] C. Bandt, B. Pompe, Permutation Entropy: A Natural Complexity Measure for Time Series, Phys. Rev. Lett. 88 (2002) 174102. doi:10.1103/PhysRevLett.88.174102.

[27] S. Charbonnier, L. Zoubek, S. Lesecq, F. Chapotot, Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging, Comput. Biol. Med. 41 (2011) 380–389. doi:10.1016/j.compbiomed.2011.04.001.

[28] A. Noviyanto, A.M. Arymurthy, Sleep stages classification based on temporal pattern recognition in neural network approach, in: 2012 Int. Jt. Conf. Neural Networks, IEEE, 2012: pp. 1–6. doi:10.1109/IJCNN.2012.6252386.

[29] N. Sriraam, B.R. Purnima, K. Uma, T.K. Padmashri, Hurst exponents based detection of wake-sleep &amp;#x2014; A pilot study, in: Int. Conf. Circuits, Commun. Control Comput., IEEE, 2014: pp. 118–121. doi:10.1109/CIMCA.2014.7057771.

[30] R. Acharya U., O. Faust, N. Kannathal, T. Chua, S. Laxminarayan, Non-linear analysis of EEG signals at various sleep stages, Comput. Methods Programs Biomed. 80 (2005) 37–45. doi:10.1016/j.cmpb.2005.06.011.

[31] K. Aboalayon, M. Faezipour, W. Almuhammadi, S. Moslehpour, Sleep Stage Classification Using EEG Signal Analysis: A Comprehensive Survey and New Investigation, Entropy. 18 (2016) 272. doi:10.3390/e18090272.

[32] X. Kong, N. Thakor, V. Goel, Characterization of EEG signal changes via Itakura distance, in: Proc. 17th Int. Conf. Eng. Med. Biol. Soc., 1995: pp. 873–874. doi:10.1109/IEMBS.1995.579247.

[33] E. Estrada, H. Nazeran, P. Nava, K. Behbehani, J. Burk, E. Lucas, EEG feature extraction for classification of sleep stages., in: Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf., IEEE, 2004: pp. 196–199. doi:10.1109/IEMBS.2004.1403125.

[34] E. Estrada, P. Nava, H. Nazeran, K. Behbehani, J. Burk, E. Lucas, Itakura Distance: A Useful Similarity Measure between EEG and EOG Signals in Computer-aided Classification of Sleep Stages., Conf. Proc. IEEE Eng. Med. Biol. Soc. 2 (2005) 1189–1192. doi:10.1109/IEMBS.2005.1616636.

[35] F. Ebrahimi, M. Mikaili, E. Estrada, H. Nazeran, Assessment of Itakura Distance as a Valuable Feature for Computer-aided Classification of Sleep Stages, in: 2007 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE, 2007: pp. 3300–3303. doi:10.1109/IEMBS.2007.4353035.

[36] F. Chapotot, G. Becq, Automated Sleep-Wake Staging Combining Robust Feature Extraction, Artificial Neural Network Classification, and Flexible Decision Rules, Int. J. Adapt. Control Signal Process. 24 (2009) 409–423. doi:10.1002/acs.1147.

[37] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, F. Chapotot, Feature Selection for Sleep/Wake Stages Classification Using Data Driven Methods, Biomed. Signal Process. Control. 2 (2007) 171–179. doi:10.1016/j.bspc.2007.05.005.

[38] L. Fraiwan, K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, H. Dickhaus, Classification of Sleep Stages Using Multi-wavelet Time Frequency Entropy and LDA, Methods Inf. Med. 49 (2010) 230–237. doi:10.3414/ME09-01-0054.

[39] R.K. Sinha, Artificial Neural Network and Wavelet Based Automated Detection of Sleep Spindles, REM Sleep and Wake States, J. Med. Syst. 32 (2008) 291–299. doi:10.1007/s10916-008-9134-z.

[40] S. Gudmundsson, T.P. Runarsson, S. Sigurdsson, Automatic Sleep Staging using Support Vector Machines with Posterior Probability Estimates, in: Int. Conf. Comput. Intell. Model. Control Autom. Int. Conf. Intell. Agents, Web Technol. Internet Commer., IEEE, 2005: pp. 366–372. doi:10.1109/CIMCA.2005.1631496.

[41] T. Sousa, A. Cruz, S. Khalighi, G. Pires, U. Nunes, A two-step automatic sleep stage classification method with dubious range detection, Comput. Biol. Med. 59 (2015) 42–53. doi:10.1016/j.compbiomed.2015.01.017.

[42] S. Güneş, K. Polat, Ş. Yosunkaya, Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting, Expert Syst. Appl. 37 (2010) 7922–7928. doi:10.1016/j.eswa.2010.04.043.

[43] M. Längkvist, L. Karlsson, A. Loutfi, Sleep Stage Classification Using Unsupervised Feature Learning, Adv. Artif. Neural Syst. 2012 (2012) 1–9. doi:10.1155/2012/107046.

[44] O. Tsinalis, P.M. Matthews, Y. Guo, Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders, Ann. Biomed. Eng. 44 (2016) 1587–1597.

doi:10.1007/s10439-015-1444-y.

[45] S. Najdi, A.A. Gharbali, J.M. Fonseca, Feature ranking and rank aggregation for automatic sleep stage classification: a comparative study, Biomed. Eng. Online. 16 (2017) 78. doi:10.1186/s12938-017-0358-3.

[46] PhysioNet, The Sleep-EDF Database [Expanded], (2015). doi:10.13026/C27C7Q.

[47] S. Khalighi, T. Sousa, J.M. Santos, U. Nunes, ISRUC-Sleep: A comprehensive public dataset for sleep researchers, Comput. Methods Programs Biomed. 124 (2016) 180–192. doi:10.1016/j.cmpb.2015.10.013.

[48] M. Obayya, F.E.Z. Abou-Chadi, Automatic classification of sleep stages using EEG records based on Fuzzy c-means (FCM) algorithm, in: 2014 31st Natl. Radio Sci. Conf., IEEE, 2014: pp. 265–272. doi:10.1109/NRSC.2014.6835085.

[49] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, Y.-H. Pan, Y.-H. Wang, Automatic Stage Scoring of Single-Channel Sleep EEG by Using Multiscale Entropy and Autoregressive Models, IEEE Trans. Instrum. Meas. 61 (2012) 1649–1657. doi:10.1109/TIM.2012.2187242.

[50] F. Ebrahimi, M. Mikaeili, E. Estrada, H. Nazeran, Automatic Sleep Stage Classification Based on EEG Signals by Using Neural Networks and Wavelet Packet Coefficients, 2008 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. 2008 (2008) 1151–1154. doi:10.1109/IEMBS.2008.4649365.

[51] B. Iser, W. Minker, G. Schmidt, Bandwidth extension of speech signals, in: Lect. Notes Electr. Eng., 2008: pp. 1–182. doi:10.1007/978-0-387-68899-2.

[52] M.M. Deza, E. Deza, Encyclopedia of distances, 2009. doi:10.1007/978-3-642-00234-2.

[53] M. Brookes, VOICEBOX: Speech Processing Toolbox for MATLAB, (2005). http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.

[54] A. Kalousis, J. Prados, M. Hilario, Stability of Feature Selection Algorithms: A Study on High-Dimensional Spaces, Knowl. Inf. Syst. 12 (2007) 95–116. doi:10.1007/s10115-006-0040-8.

[55] K. Kira, L.A. Rendell, The Feature Selection Problem: Traditional Methods and a New Algorithm, in: W.R. Swartout (Ed.), Proc. 10th Natl. Conf. Artif. Intell. San Jose, CA, July 12-16, 1992., 1992: pp. 129–134. http://www.aaai.org/Library/AAAI/1992/aaai92-020.php.

[56] M. Robnik-Šikonja, I. Kononenko, Theoretical and Empirical Analysis of ReliefF and RReliefF, Mach. Learn. 53 (2003) 23–69. doi:10.1023/A:1025667309714.

[57] C. Ding, H. Peng, Minimum Redundancy Feature Selection from Microarray Gene Expression Data, in: Comput. Syst. Bioinformatics. CSB2003. Proc. 2003 IEEE Bioinforma. Conf. CSB2003, IEEE Comput. Soc, 2005: pp. 523–528. doi:10.1109/CSB.2003.1227396.

[58] Q. Gu, Z. Li, J. Han, Generalized Fisher Score for Feature Selection, in: Proc. Int. Conf. Uncertain. Artif. Intell., 2011. http://arxiv.org/abs/1202.3725.

[59] Huan Liu, R. Setiono, Chi2: Feature Selection and Discretization of Numeric Attributes, in: Proc. 7th IEEE Int. Conf. Tools with Artif. Intell., IEEE Comput. Soc. Press, 1995: pp. 388–391. doi:10.1109/TAI.1995.479783.

[60] J.R. Quinlan, C4.5: Programs for Machine Learning, (1993).

[61] K. Benabdeslem, Y. Bennani, Dendogram based SVM for multi-class classification, in: 28th Int. Conf. Inf. Technol. Interfaces, 2006., IEEE, 2006: pp. 173–178. doi:10.1109/ITI.2006.1708473.

[62] F. Takahashi, S. Abe, Decision-tree-based multiclass support vector machines, Proc. 9th Int. Conf. Neural Inf. Process. 2002. ICONIP ’02. 3 (2002) 1418–1422 vol.3. doi:10.1109/ICONIP.2002.1202854.

[63] G. Madzarov, D. Gjorgjevikj, I. Chorbev, A Multi-class SVM Classifier Utilizing Binary Decision Tree, Informatica. 33 (2009). http://www.informatica.si/index.php/informatica/article/view/241 (accessed July 6, 2016).

[64] M. Bala, R.K. Agrawal, Optimal Decision Tree Based Multi-class Support Vector Machine, Informatica. 35 (2011).

[65] L. Duckstein, S. Opricovic, Multiobjective optimization in river basin development, Water Resour. Res. 16 (1980) 14–20. doi:10.1029/WR016i001p00014.

[66] S. Opricovic, G.H. Tzeng, Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS, Eur. J. Oper. Res. 156 (2004) 445–455. doi:10.1016/S0377-2217(03)00020-1.

[67] G. Sharma, J. Martin, MATLAB®: A Language for Parallel Computing, Int. J. Parallel Program. 37 (2009) 3–36. doi:10.1007/s10766-008-0082-5.

[68] A. Krakovská, K. Mezeiová, Automatic sleep scoring: A search for an optimal combination of measures, Artif. Intell. Med. 53 (2011) 25–33. doi:10.1016/j.artmed.2011.06.004.

[69] P. Anderer, G. Gruber, S. Parapatics, M. Woertz, T. Miazhynskaia, G. Klösch, B. Saletu, J.

Zeitlhofer, M.J. Barbanoj, H. Danker-Hopfe, S.-L. Himanen, B. Kemp, T. Penzel, M. Grözinger, D. Kunz, P. Rappelsberger, A. Schlögl, G. Dorffner, An E-Health Solution for Automatic Sleep Classification according to Rechtschaffen and Kales: Validation Study of the Somnolyzer 24 × 7 Utilizing the Siesta Database, Neuropsychobiology. 51 (2005) 115–133. doi:10.1159/000085205.

[70] S. Khalighi, T. Sousa, G. Pires, U. Nunes, Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels, Expert Syst. Appl. 40 (2013) 7046–7059. doi:10.1016/j.eswa.2013.06.023.

[71] M. Rempe, W. Clegern, J. Wisor, An automated sleep-state classification algorithm for quantifying sleep timing and sleep-dependent dynamics of electroencephalographic and cerebral metabolic parameters, Nat. Sci. Sleep. 7 (2015) 85. doi:10.2147/NSS.S84548.