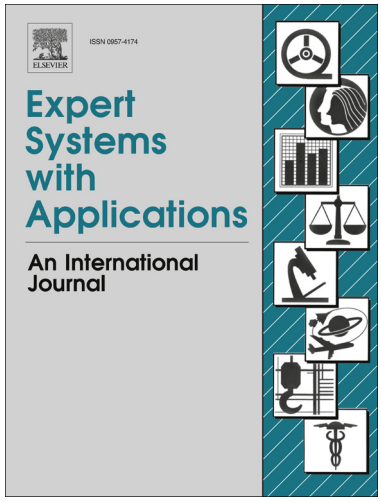

Automatic Sleep Staging: A Computer Assisted Approach for Optimal Combination of Features and Polysomnographic Channels

Sirvan Khalighi, Teresa Sousa, Gabriel Pires, Urbano Nunes

PII: S0957-4174(13)00403-X
DOI: http://dx.doi.org/10.1016/j.eswa.2013.06.023
Reference: ESWA 8636

To appear in: *Expert Systems with Applications*

Please cite this article as: Khalighi, S., Sousa, T., Pires, G., Nunes, U., Automatic Sleep Staging: A Computer Assisted Approach for Optimal Combination of Features and Polysomnographic Channels, *Expert Systems with Applications* (2013), doi: http://dx.doi.org/10.1016/j.eswa.2013.06.023

Automatic Sleep Staging: A Computer Assisted Approach for Optimal Combination of Features and Polysomnographic Channels

Sirvan Khalighi, Teresa Sousa, Gabriel Pires and Urbano Nunes

Institute of Systems and Robotics (ISR-UC), Department of Electrical and Computer Engineering University of Coimbra, Coimbra, Portugal

Corresponding author: Sirvan Khalighi, E-mail: skhalighi@isr.uc.pt,

Abstract

To improve applicability of automatic sleep staging an efficient subject-independent method is proposed with application in sleep-wake detection and in multiclass sleep staging (awake, non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep). In turn, NREM is further divided into three stages denoted here by N1, N2, and N3. To assess the method, polysomnographic (PSG) records of 40 patients from our ISRUC-Sleep dataset, which was scored by an expert clinician in the central hospital of Coimbra, are used. To find the best combination of PSG signals for automatic sleep staging, six electroencephalographic (EEG), two electrooculographic (EOG), and one electromyographic (EMG) channels are analyzed. An extensive set of feature extraction techniques are applied, covering temporal, frequency and time-frequency domains. The maximum overlap wavelet transform (MODWT), a shift invariant transform, was used to extract the features in time-frequency domain. The extracted feature set is transformed and normalized to reduce the effect of extreme values of features. The most discriminative features are selected through a two-step method composed by a manual selection step based on features' histogram analysis followed by an automatic feature selector. The selected feature set is classified using support vector machines (SVMs). The system achieved the best performance by combining 6 channels (C3, C4, O1, left EOG (LOC), right EOG (ROC) and chin EMG (X1)) for sleep-wake detection, and 9 channels (C3, C4, O1, O2, F3, F4, LOC, ROC, X1) for multiclass sleep staging.

Keywords: Automatic Sleep Staging, The maximum overlap discrete wavelet transform, Polysomnographic signals, Features selection, Sleep dataset.

This dataset, as well as MATLAB code of the proposed algorithm will be made publicly available.

1. Introduction

Sleep is an active and regulated process with an essential restorative function for physical and mental health [1]. Sleep disorders have an important effect on the health and quality of life. Sleep staging is an essential part of the diagnostic process in the assessment of sleep disorders such as Sleep Apnea Syndrome (SAS) [2]. Therefore, monitoring, scoring and detecting abnormal changes of sleep pattern through whole night sleep recordings have consistently been an important research topic. Scoring of sleep stages was done on the basis of Rechtschaffen and Kales standard (R&K) until recent dates [3]. The American Academy of Sleep Medicine (AASM) determined new criteria in the scoring of sleep based on the R&K rules. In adults, sleep-wake cycle is categorized in awake, non-rapid eye movement (NREM) and rapid eye movement (REM) sleep stages. NREM sleep is further divided into three stages: N1, N2 and N3 [4], the last of which is also called delta sleep or slow wave sleep(SWS). Moreover, the sleep stages during a night sleep, proceeds in cycles of NREM and REM, each cycle normally being N1 → N2 → N3 → N2 → REM. The cycles typically happen 4 to 6 times during whole night sleep [5]. The AASM rules define some characteristics for each sleep stages according to the amplitude, frequency and shape of the polysomnographic (PSG) signals (see Table I). Manual visual sleep scoring by highly trained human experts is a very time consuming task and normally may require hours to score the PSG recording of a whole night. Visual interpretation of PSG records based on AASM uses fixed epoch duration 30 seconds, and allows for the recognition of different sleep-wake stages (Fig. 1). It is also a somewhat subjective procedure in which the concordance between the results of visual scoring obtained by experts can vary greatly. Accordingly,

Table I
Summary of EEG,EOG and EMG patterns for different sleep stages

Stages	**EEG**					**EOG**	**EMG**
	Delta (< 4 Hz)	*Theta (4 - 7 Hz)*	*Alpha (8 - 13 Hz)*	*Beta (> 13 Hz)*	*Other EEG patterns*		
AWAKE			x	x		0.5-2 Hz	Variable amplitude but usually higher than during sleep stages
N1		x	x		Vertex waves	Slow eye movement	Lower amplitude than in stage awake
N2		x			K-complexes; Sleep spindles	Usually no eye movement, but slow eye movements may persist	Lower amplitude than in stage awake and may be as low as in stage REM
N3	x				Sleep spindles may persist	Eye movements are not typically seen	Lower amplitude than in stage N2 and sometimes as lower as in stage REM
REM		x	x		Sawtooth waves	Rapid eye movement	Low chin EMG tone; usually the lowest level of entire recording

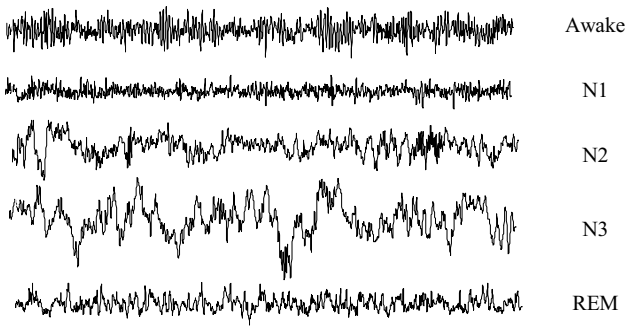

Fig. 1. EEG pattern of different sleep stages

an efficient automatic sleep scoring may save time and provide objective assessment of sleep, independent of subjective interpretation of experts.

Several studies have reported the development of automatic sleep stage classification (ASSC) methods based on PSG records, namely electroencephalographic (EEG) records, sometimes in combination with electrooculographic (EOG) and electromyographic (EMG) records collected from human individuals using noninvasive surface electrodes. Some of the most important published works are summarized in Table II. Most of the works employed frequency, or time-frequency domains' features such as discrete Wavelet transform (DWT), Hilbert Huang transform (HHT), and fast Fourier transform (FFT) [1, 6-8]. Different parametric and nonparametric methods have been applied in the classification process such as random forest classifiers, artificial neural networks (ANN), fuzzy logic, the nearest neighbour, linear discriminant analysis (LDA,) support vector machine (SVM) and kernel logistic regression (KLR) [6-12]. Classification accuracies vary widely among the ASSC methods reported in scientific literature. Rigorous comparisons between the reported systems cannot be done since they differ in recording conditions and validation procedures. State-of-the-art results summarized in Table II show agreement level with the manual scores ranging from 55% to 85%.

The contributions of this work are fourfold:

1. A new ASSC method has been developed, aiming to improve sleep stage classification accuracy in two applications: sleep-wake detection and multiclass sleep staging classification. The effectiveness of our approach is demonstrated through a series of experiments involving PSG data from our extensive dataset of 40 different subjects with confirmed or only suspicious sleep disorders which was collected in central hospital of Coimbra.
2. A new time-frequency based feature extraction method, for ASSC is proposed. To decompose the PSG signals in different resolutions, the

Table II
Performance results to the multiclass sleep staging.

ASSC approaches	Sleep stages	Nature of feature	Matching process	Subjects/ Channels	Quality evaluation
Zoubek *et al.*[1]	W, NREM-S1, NREM-S2, SWS, REM	30 sec. epoch; 10 features, EEG: RP delta, theta, alpha, sigma, beta (FT coefficients), 75th percentile; EMG: entropy; EOG: entropy, kurtosis number and SD.	Neural Network BP MLP	47 recordings, EEG, EMG	71% (EEG only), 80% (EEG, EOG and EMG): W: 84.57%, S1:64.56%, S2:85.55%, SWS:92.90%, REM: 72.81%.
Jo *et al.*[6]	Wakefulness (WA), shallow sleep (SS), deep sleep (DS), and rapid eye movement (REM)	30 sec. epoch; Fast Fourier transform (FFT) with Hamming window; power spectra; Relative powers(RP)	Fuzzy classifier and a genetic algorithm (GA).	4 recording, single EEG(C3-A2)	84.60%
Tang *et al.*[7]	Awake, NREM-S1, NREM- S2, NREM-S3, NREM- S4, REM	30 sec. epoch; HHT, Wavelet Transform, Autoregressive model	SVM	6 recordings, EEG(C3-A2), EMG and EOG	Wavelet: 77.9% HHT: 77.6%
Fraiwan *et al.* [8]	W, sleep N1, sleep N2, sleep N3, REM	30 sec. epoch; Choi–Williams distribution (CWD), Continuous wavelet transform (CWT), and HHT and Renyi's entropy measures	Random forest classifier	16 recording, Single EEG (C3-A1)	Accuracy 83%; and kappa coefficient of 0.76.
Gunes *et al.*[9]	W, NREM-N1, NREM-N2, NREM-N3, REM	30 sec. epoch; 129 features: Welch spectral analysis; k-means clustering based feature weighting (KMCFW)	K-nearest neighbour (KNN) and C4.5 decision tree	4 recording, EEG	55.88% by k-NN; the weighted sleep stages with KMCFW has been recognized with 82.15% success
Fraiwan *et al.*[13]	W, NREM-S1, NREM-S2, NREM-S3, NREM- S4, REM	30 sec. epoch; Entropy on CWT, used three different mother wavelets	Linear discriminant analysis (LDA)	32 recording from MIT-BIH, Single EEG	Accuracy 84%, kappa coefficient 0.78.
Estévez *et al.*[14]	Awake, NREM-S1, NREM- S2, NREM-S3, NREM- S4, REM	Amplitude of EOG, EMG, short time FFT, power Spectral density on FFT	Continuous fuzzy reasoning scheme	33 recordings, EEG(C3-A2 and C4-A1), EMG and EOG	W: 34%, N1: 43%, N2: 51%, N3: 82%, REM: 82%,
Helland *et al.*[15]	W, NREM-N1, NREM-N2, NREM-N3, REM	30 sec. epoch; power (P) and beta/delta, alpha/delta, theta/delta, beta/theta, alpha/theta, beta/alpha, beta/P, alpha/P, theta/P, and delta/P; heart rate variability (HRV) parameters	LDA	10 recording, EEG, ECG, and respiratory signals	90% Just EEG; By including EMG and respiratory signals 93%, Agreement with visual 61%
Tagluk *et al.*[16]	REM, NREM-S1, NREM- S2, NREM-S3, NREM- S4	5 sec. epoch; 5 features	Neural Network BP MLP, RUM with momentum	21 recordings, EEG(C3-A2), EMG and EOG	W: 70.5%, NREM :82.6% REM: 38.3%,

Chapotot and Bequ [17]	W,NREM-N1, shallow NREM-N2, deep NREM-N3, REM, MT	20 sec. epoch with small subset of 2 sec. epochs; 16 features: Shannon entropy, Hjorth activity, mobility and complexity, Hurst exponent, spectral edge frequency 95%, RP	Neural Network BP MLP, and flexible decision rules	48 recordings, EEG, EMG	W: 34%, N1: 43%, N2: 51%, N3: 82%, REM: 82%, MT: 13%

maximum overlap discrete wavelet transform (MODWT), which is shift invariant transform, is employed. Moreover, since temporal and frequency based features represent other aspects of the signals, several temporal and frequency based methods for feature extraction have been investigated.

3. Some works such as [8, 13] used just one or more EEG channels, whereas others [1, 7, 14] used EEG channels in combination with EOG and EMG channels. Therefore, to reduce the computational cost and improve classification performance, a systematic analysis for finding the best combination of EEG, EOG and EMG channels, for both application sleep-wake detection and multiclass sleep staging, is performed.
4. Current automatic feature selectors are dependent to the classifiers; moreover, they are affected by extreme values in feature vectors. Therefore to find the most discriminative features for sleep-wake detection and multiclass sleep staging a two-step feature selector is applied on the transformed and normalized feature vectors. This two-step algorithm is composed by a manual selection followed by an automatic selector. For the second part of the algorithm, six different feature selectors are investigated.

2. Subjects and Signals under Study

Data from all-night PSG records, each with duration around 8 hours (acquired by a SomnoStar Pro; Viasys SensorMedics, a multi-channel ambulatory recording device), were provided by the Laboratory of Sleep from central hospital of Coimbra. All EEG, EOG, and EMG (chin) recordings were performed with a sampling rate of 200 Hz and stored into computer files using the standard EDF data format [18]. The international 10-20 standard electrode placement system [19] was used for EEG recording. The PSG is composed by signals from the following 19 channels:

- 6 EEG (F3, C3, O1, F4, C4 and O2);
- 2 EOG, right and left (ROC and LOC);

- 1 electrocardiographic (ECG);
- 2 types of EMG (one m. submentalis – chin EMG (X1) – and two m. tibialis – legs EMG);
- 1 snore (derived);
- 2 airflow (pressure based);
- 2 abdominal effort;
- 1 pulse oximetry (SaO_2);
- 1 body position (BPOS);
- The ground and reference were placed in the right earlobe.

All recordings were segmented into epochs of 30 seconds and visually labelled by an expert according to the guidelines of AASM [4], with the stages: awake, NREM (N1, N2, N3) and REM sleep. Our ISRUC-Sleep dataset comprises data from forty adult subjects (detailed in Table VIII), twenty six males and fourteen females with ages between 22 and 85 years old (mean = 54.35 years; STD = 16.37 years), with suspicious of sleep disorders, most of them with detected sleep apnea events. The subjects can be medicated but they can breathe without machine help. Six EEG, two EOG channels and one EMG channel were used in our evaluation: F3-A2, C3-A2, O1-A2, F4-A1, C4-A1, O2-A1, ROC-A1, LOC-A2, and X1(EMG) for all the subjects. To validate the results, in each new test, the initial population set is divided in two independent groups based on Leave-one subject-out cross-validation (LOOCV) strategy. The training set is used to obtain the most discriminative feature subset and training model created by a classifier. On the other hand, the test set (test subject) is used to assess the proposed method.

3. Methodology and Algorithm Description

The proposed system is organized in various interoperating parts as detailed in the Fig. 2: preprocessing, feature extraction, feature transformation and normalization, feature selection, and classification.

3.1. Preprocessing and Feature Extraction

The recorded signals are filtered to eliminate noise and undesired background EMG, by using a notch filter at 50 Hz and a band-pass Butterworth filter with lower cut-off of 0.5 Hz and higher cut-off of 45 Hz. The signals were segmented in 30 seconds epochs.

PSG is traditionally analyzed in the frequency domain, since each sleep stage is characterized by a specific pattern of frequency contents. However, further useful information can be extracted from temporal analysis of PSG signals given the nonstationary PSG signals, time-frequency transformations like wavelets are

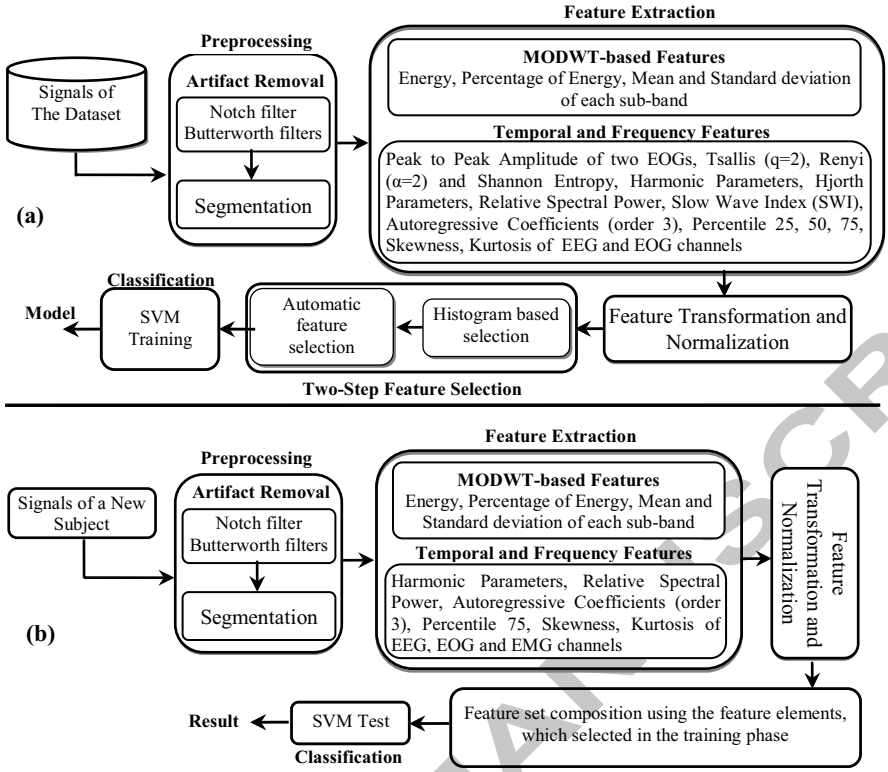

Fig. 2. System architecture; (a) training phase, (b) test phase.

very useful. Thus, after preprocessing, some features are extracted using several methods in the time-frequency, temporal and frequency domains.

3.1.1. The Maximum Overlap Discrete Wavelet Based Features

Wavelet transform acts like a mathematical microscope zooming into small scales to reveal compactly spaced events in time and zooming out into large scales to exhibit the global waveform patterns [20]. The discrete wavelet transform (DWT) generates coefficients, which are local in time and frequency and represent the energy distribution of the signals. Therefore, signals can be reconstructed as a linear combination of the wavelet functions weighted by the wavelet coefficients. The maximum overlap discrete wavelet transform (MODWT) [21] is a DWT in which the operation of sub-sampling from an output filter is omitted. By giving up of the orthogonality property, the MODWT gains new features; although losing efficiency in computation, this transform does not have any restriction on the sample size and it is shift invariant. As a result, in the MODWT, the wavelet and scaling coefficients must be rescaled to retain the variance preserving property of the DWT. Although the components of MODWT are not mutually orthogonal, their sum is equal to the original time series. Additionally, the detail and smooth coefficients of a MODWT are

Table III
Frequencies corresponding to different decomposition levels

Decomposition	Frequency range (Hz)
D1	25–50
D2	12.5–25
D3	6.25–12.5
D4	3.125–6.25
D5	0–3.125

associated with zero phase filters. This means that temporal events and patterns in the original signal are meaningfully aligned with the features in the multi resolution analysis. Furthermore, the MODWT is invariant to circularly shifting the original time series. Hence, shifting the time series by an integer unit will shift the wavelet and scale coefficients by the same amount. This property does not hold for the DWT because of the sub-sampling involved in the filtering process. In addition, the MODWT does not induce the phase shifts within the component series. The MODWT variance estimator is also preferred because it has been shown to be asymptotically more efficient than an estimator based on the DWT [22]. In this study a MODWT of depth 6 with Daubechies order four (db4) is applied to every 30 second epochs with a sampling rate of 200 Hz. As shown in Table III, the frequency ranges are broken down in a decomposition of D1-D5, which almost correspond to δ range (<4 Hz), θ range (4–8 Hz), α range (8–13 Hz) and β range (13–30 Hz). Finally, a set of statistical MODWT-based features are extracted to represent the time–frequency distribution of the EEG, EOG and EMG signals.

Energy and Percentage of Energy: Parseval's theorem is employed to extract the distribution of energy of the signals. According to Parseval's theorem, the energy of the distorted signal can be partitioned at different resolution levels. Mathematically it can be presented as:

$$E_i = \sum_{j=1}^{N} |D_{ij}|^2, i = 1, \dots . l \quad (1)$$

where i =1,.....,l denote the MODWT decomposition level. E_i is energy at decomposition level i, N is the number of the coefficients at each decomposition level and D_{ij} is value of a coefficient j at decomposition level i :

$$PE_i = E_i / \sum_{j=1}^{l} (E_j) \quad (2)$$

PE_i is Percentage energy at decomposition level i [23].

Mean and standard deviation of each sub-band: In order to reduce the dimensionality of the extracted feature vectors, mean (M_i) and standard deviation Std_i at decomposition level i are used.

$$M_i = \sum_{j=1}^{N} (D_{ij}) / N \quad (3)$$

$$Std_i = \sqrt{\frac{\sum_{j=1}^{N}(D_{ij}-M_i)^2}{N-1}} \quad (4)$$

3.1.2. Frequency and Temporal Features

Due to the importance of spectral and temporal analysis, some features are extracted in these domains. The following features are suggested in [1, 7, 24, 25].

Peak to Peak Amplitude: Peak-to-peak amplitude $P(X)$ is calculated by

$$P(X) = max(X) - min(X) \quad (5)$$

where $X = \{x_1, x_2 \dots, x_N\}$ denotes set of signal amplitudes.

Entropy: The entropy gives a measure of signal disorder and can provide relevant information in the detection of some signal disturbs. Shannon entropy [26] is computed from histogram of the PSG samples, where $p_1, p_2, \dots, p_n$ are a series of events; $p_i = n_i/N$, where N is the number of samples within the signal X, and n_i is the number of samples within the i^{th} bin. Shannon entropy H is defined as

$$H(P) = -K\sum_{i=1}^{N} p_i ln p_i \quad (6)$$

where K is a positive constant.

Extensions of Shannon's original work have resulted in many alternative measures of information or entropy. Renyi [27] was able to extend Shannon entropy to a continuous family of entropy measures that obey

$$H_q(P) = \frac{1}{1-q} ln \sum_{i=1}^{N} p_i^q \quad (7)$$

The Renyi entropy tends to Shannon entropy as $q \to 1$.

Furthermore, recently Tsallis entropy has proposed the use of the same quantity as a physical entropy measure which has some provoked considerable controvers [28]. Tsallis defined his entropy as [29] :

$$T_q = \frac{1}{q-1}\left(1 - \sum_{i=1}^{n} p_q^i\right) \quad (8)$$

Relative Spectral Power (RSP): Spectral analysis provides some of the most important features. For each signal *X*, an FFT squared modulus estimator was applied to estimate the power spectral density (PSD). The spectrum is divided into five frequency sub-bands as represented in Table III. For each frequency sub-band, the RSP is computed. This parameter is given by the ratio between the sub-band spectral power (BSP) and the total spectral power, i.e., the sum of all five BSP sub-band [30]. Moreover, the spectral bands *Delta*, *Theta* and *Alpha* can be highlighted over slow wave bands by means of slow wave indexes defined by the following ratios:

Table IV
Spectral sub-bands used in PSD computation

Bands	Sub-bands	Bandwidth f_L-f_H (Hz)
Delta	Delta 1	0.5-2.0
	Delta 2	2.0-4.0
Theta	Theta 1	4.0-6.0
	Theta 2	6.0-8.0
Alpha	Alpha 1	8.0-10.0
	Alpha 2	10.0-12.0
Sigma	Sigma 1	12.0-14.0
	Sigma 2	14.0-16.0
Beta	Beta 1	16.0-25.0
	Beta 2	25.0-35.0

$$DSI = \frac{BSP_{Delta}}{BSP_{Theta}+BSP_{Alpha}} \tag{9}$$

$$TSI = \frac{BSP_{Theta}}{BSP_{Delta}+BSP_{Alpha}} \tag{10}$$

$$ASI = \frac{BSP_{Alpha}}{BSP_{Delta}+BSP_{Theta}} \tag{11}$$

where TSI, ASI [31] and DSI stand for theta-slow-wave index, alpha-slow-wave index and delta-slow-wave index, respectively.

Harmonic Parameters: Harmonic Parameters of the PSG signals include three parameters: the center frequency (f_c) (12), the bandwidth (f_σ) (13) and the spectral value at center frequency (S_{fc}) (14). These parameters are defined as follows [7]:

$$f_c = \sum_{f_L}^{f_H} f p_{xx}(f) / \sum_{f_L}^{f_H} p_{xx}(f) \tag{12}$$

$$f_\sigma = \sqrt{\sum_{f_L}^{f_H} (f - f_c)^2 f p_{xx}(f) / \sum_{f_L}^{f_H} p_{xx}(f)} \tag{13}$$

$$S_{f_c} = p_{xx}(f_c) \tag{14}$$

where $p_{xx}(f)$ denotes the PSD, which is calculated for the frequency bands f_L-f_H (see Table IV). These parameters allow the analysis of a specific band in the EEG spectrum.

Hjorth Parameters: The Hjorth parameters provide dynamic temporal information of the PSG signals. The Activity, Mobility and Complexity parameters are computed from the variance X, ($var(X)$) and the first and second derivatives X', X''according to [32]:

$$Activity = var(X) \tag{15}$$

$$Mobility = \sqrt{var(X')/var(X)} \tag{16}$$

$$Complexity = \sqrt{var(X'') \times var(X)/var(X')^2} \tag{17}$$

Skewness and Kurtosis: The skewness describes a measure of symmetry, or more precisely, the lack of symmetry of a distribution. Skewness of a signal X with N samples x_i is defined as

$$Skewness = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i}{\sigma}\right)^3 \quad (18)$$

As formulated in (19) the kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution.

$$Kurtosis = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i}{\sigma}\right)^4\right] - 3 \quad (19)$$

Autoregressive Coefficients: Autoregressive (AR) model is a representation of a time series such that it specifies that the output variable depends linearly on its own previous values. An AR process is defined by

$$x_i = \sum_{j=1}^{N} a_j x_{i-j} + \varepsilon_i \quad (20)$$

where a_j are the autoregression coefficients, x_i is the series under investigation, which is a linear combination of its N past values and a purely random process ε_i. The noise term or residue, epsilon in the above, is almost always assumed to be Gaussian white noise [30].

Percentile 25, 50, 75: The percentile analysis provides some information about the amplitude of the signal and might be useful in discerning certain sleep stages [1]. The 25th, 50th and 75th percentile of the signal distribution is defined as

$$Percintile_P(X) = (PN)/100 \quad (21)$$

where N is the number of samples x_i of the measured signal X and $P \in \{25,50,75\}$

Table V
Transformation methods [33]

#	Transformation	#	Transformation	#	Transformation
T1	$1/\sqrt{x}$	T4	Log(x)	T7	Log(x/(1-x))
T2	$\sqrt[3]{x}$	T5	Log(1+x)		
T3	$\sqrt{x}$	T6	arcsin($\sqrt{x}$)		

3.2. Feature Transformation and Normalization

The extracted features are transformed and normalized in order to reduce the influence of extreme values. The transformation methods applied to each feature are described in Table V [33]. It was verified that some of those transformations improved the classification results. After a thorough experimental evaluation of each transform operator over extracted features, it was verified that the best classification results were attained with the transform

$$\boldsymbol{X} = \arcsin\left(\sqrt{\boldsymbol{Y}}\right) \quad (22)$$

where $\boldsymbol{Y}$ denotes the feature matrix, and

$$\boldsymbol{X} = \{x_{ij}; i = 1,2,\dots,N \text{ and } j = 1,2,\dots,M\} \quad (23)$$

is the transformed feature matrix, where N and M denote the number of subjects and the number of features, respectively. Thereby this transform was adopted in the overall sleep staging system.

To avoid features in greater numeric ranges dominating those in smaller numeric ranges, as well as numerical difficulties during classification; each feature of the transformed matrix $\boldsymbol{X}$ is independently normalized to the [0, 1] range by applying

$$\bar{x}_{ij} = x_{ij}/\left(max(\boldsymbol{x}_j) - min(\boldsymbol{x}_j)\right) \quad (24)$$

where $\boldsymbol{x}_j$ is a vector of each independent feature.

3.3. Feature Selection

To reduce the dimension of the features vector and to find the most discriminative features, a two-step method that consists on a filtering and a wrapper phases is proposed: firstly, as detailed in Algorithm 1, the less discriminative feature-types are removed. In fact, by investigation on the feature distribution's histogram and corresponding hypnogram during a whole night

Algorithm 1: Two-step Feature Selection method.

Feature Selection (*ExtractedFeatureSet***)**

Featurevector = {F_1, F_2, …, F_N} , F_i = {y_1, y_2, …, y_M},

$F_{i\,Sleep} = \{F_{i1}, F_{i2}, .., F_{iz}\}$.

% N: number of feature type, M: number of the feature elements,

% Z: number of sleep epochs

Step 1

1. Set SelectedFeatureType = { }, d =1.

% initializes preliminary set of features

While (d <= N) do:

a. Comparison of feature distribution $F_{d\,Sleep}$ with the corresponding hypnogram

b. If ($F_{d\,Sleep}$ has different distribution for sleep stages)

i. Add F_d to SelectedFeatureType.

[End of if structure]

[End of *While* structure]

Step 2

SelectedFeatureType= {y_{11}, y_{12}, …, y_{1M}, y_{21}, y_{22}, …, y_{2M}, ..., y_{k1},y_{k2}, …, y_{kM}}

% y_{ij}: an element of feature set

2. Initialize

a. Z=1, SelectedElement= {y_{11}},

b. MaxPerformance= performance ({y_{11}}).

While (Z<= length (SelectedFeatureType)) do:

i. Add y_{ij} to SelectedElement.

ii. FinalSel= AutomaticFeatureSelection (SelectedElement)

iii. If (performance(finalSel)> MaxPerformance)

iv. Maxperformance=performance(FinalSel)

v. Update SelectedElement to FinalSel

[End of *While* structure]

Return SelectedElemnts

[End of Feature Selection]

sleep, the features with a higher discriminative histogram are selected (see Fig. 10). Then, in the second step, to select the best elements of each feature-type, resulted feature vector is fed into an automatic feature selector. Feature selectors are highly dependent to defined objective function, and we are going to find the most discriminative features for ASSC independently than selector/classifiers. Therefore, to find the most discriminative feature-elements six different features selector were considered and their results were compared. Six different strategies for feature selection are described in the sequel.

Minimal-Redundancy and Maximal-Relevance: The mRMR method uses the mutual information between a feature and a class to infer its relevance for the class. The mutual information of two random variables measures the mutual dependence between them [34]. Maximal Relevance is to search a feature set S satisfying:

$$\max D(S,c), \qquad D = \frac{1}{S}\sum_{x_i \in S} I(x_i, c) \tag{25}$$

where $I(x_i; c)$ means the mutual information between feature x_i and class c. mRMR also uses the mutual information between features as redundancy of each feature. The minimal redundancy feature set R can be determined under condition

$$\min R(S), \qquad R = \frac{1}{|S|^2}\sum_{x_i, x_j \in S} I(x_i, x_j) \tag{26}$$

where $I(x_i, x_j)$ indicates the mutual information between features x_i and x_j. The "Minimal-Redundancy and Maximal-Relevance" (mRMR) criterion combines measures (25) and (26) as follows:

$$\max \varphi(D,R), \qquad \varphi = D - R \tag{27}$$

Sequential Floating Feature-Selection Approaches: Sequential forward selection (SFS) [35], which is the simplest from the sequential strategies, is a greedy search algorithm that determines iteratively an optimal subset of features by adding one feature per iteration, if it increases a chosen objective function. Sequential backward selection (SBS) [35] is similar to SFS but works in the opposite direction, i.e., it starts with the superset of all the features and sequentially removes one feature if it increases the value of the objective function.

The main drawback of these sequential approaches is that they gravitate toward local minima due to the inability to reevaluate the usefulness of features that were previously added or discarded, i.e., once a feature is added to or removed from the final set of features, it cannot be changed. Therefore, two expansions for SFS and SBS algorithms were proposed [36]. The sequential forward floating selection (SFFS) [36] finds an optimum subset by insertions

(i.e., by appending a new feature to the subset of previously selected features) and deletions (i.e., by discarding a feature from the subset of already selected features) of selected features by the SFS algorithm. The sequential backward floating selection (SBFS) [36] is similar to SFFS but works in the opposite direction; it finds an optimum subset of features by insertions (i.e., by appending an already deleted feature to the subset of selected features) and deletions (i.e., by discarding a feature from the subset of already selected features) in the SBS algorithm.

Differential Evolution Feature Selection (DEFS): DEFS approach uses a combination of differential evolution (DE) optimization method and a repair mechanism based on feature distribution measures. This method, utilizes the DE float number optimizer in the combinatorial optimization problem of feature selection. In order to make the solutions generated by the float optimizer suitable for feature selection, a roulette wheel structure is constructed and supplied with the probabilities of features distribution. These probabilities are constructed during iterations by identifying the features that contribute to the most promising solutions [37].

3.4. Classification

As classifier an SVM is applied [38]. Furthermore, LDA, Naïve Bayes (NB), and AdaBoost are used to compare the efficiency of the system.

4. Performance Assessment

The performance of the proposed algorithm was assessed using the subjects of ISRUC-Sleep dataset detailed in Section 2. Two types of experiments have been carried out: sleep-wake detection and multiclass sleep staging. In order to verify reliability of the results, all the assessments were determined by using leave-one subject-out cross-validation (LOOCV). In our experiments, a fourth order Daubechies with MODWT decomposition was adopted. Also mRMR algorithm [34] and Libsvm toolbox [39] with sigmoid kernel were used in the second phase of feature selector and classification phases, respectively. The sigmoid degree and C parameter of SVM were set to 0.13 and 1.25 respectively, as they produced the best empirical results. In order to characterize the performance of the method some well-known measures such as accuracy, receiver operating characteristic (ROC), balanced error rate (BER), sensitivity, specificity and confidence interval 95% were used. In particular, F-measure or balanced F-score is a weighted average of *precision* and *recall* where *precision*

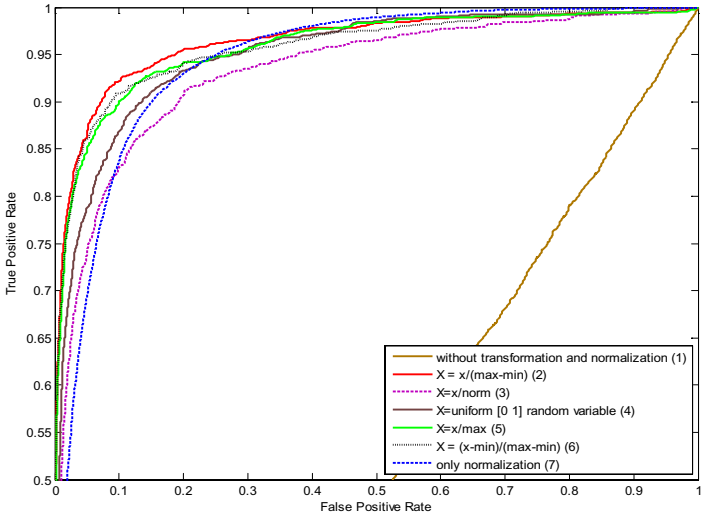

Fig. 3. ROC curves corresponding to (1) without any transformation and any normalization; (2) with normalization $\bar{x}_{ij} = x_{ij}/(max(x_i) - min(x_i))$; (3) with normalization from (2) over the transformed features by $X = \text{Arcsin}(\sqrt{Y})$, (4) and (5) normalizations *x=(x-min)/(max-min)* and *x= uniform [0 1] random variable* with the same transformation of (3).

is the fraction of retrieved instances that are relevant and *recall* is the fraction of relevant instances that are retrieved.

4.1. Evaluation of feature transformation and normalization

ROC curves related to the application of transformation and normalization methods (see Section 3.2) on extracted features are provided in Fig. 3. As it is shown, the best result was obtained by a combination of transformation *arcsin*($\sqrt{x}$) and normalization $\bar{x}_{ij} = x_{ij}/(max(x_j) - min(x_j))$. Furthermore, the performance of the system was remarkably improved when transformation and normalization operators were applied over all features. It confirms that transformation and normalization have an important effect in selection of the most discriminative features.

4.2. Evaluation of different number of features

In order to determine the best number of features in sleep-wake detection and multiclass sleep staging, a grid search was carried out over results obtained with the two-step feature selector (with mRMR) and SVM classifier. As shown in Fig. 4 and Fig. 5, the lowest average BER values occur for 147 (average BER=10.34) and 326 (average BER=15.32) features for sleep-wake and multiclass sleep staging, respectively. Nevertheless, for both cases, above 100 features the BER values do not improve significantly, e.g., in multiclass sleep staging, the BER value corresponding to 110 features is nearly similar to BER of 326 features, which performs the best result.

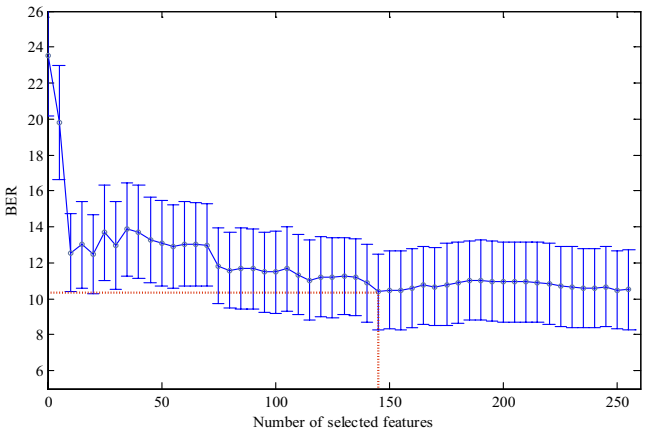

Fig. 4. Balanced Error Rate (BER) and standard deviation values corresponding to different number of selected features for sleep-awake detection.

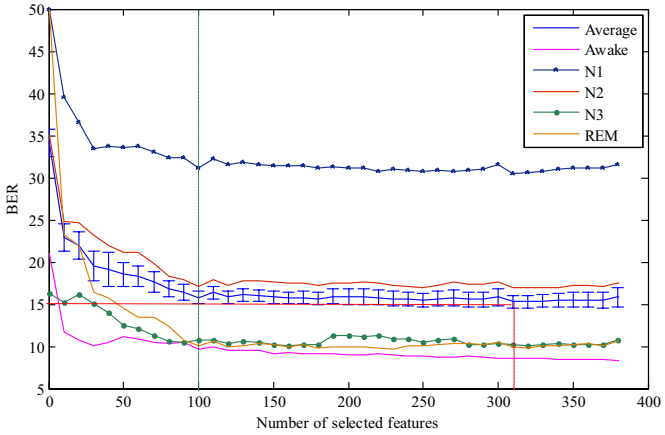

Fig. 5. Balanced Error Rate (BER) and standard deviation values corresponding to different number of selected features for: multiclass sleep staging; (1) average; (2) awake stage; (3) sleep stages N1; (4) N2; (5)N3; and (6)REM.

Table VI

Algorithm performance in Sleep-Awake detection with different channels combination

Channels	BNF	CI	AUC	ACC	BER	F-me	SEN	SPE
C3	20	0,051	81,80	84,51	18,201	72,141	72,14	91,46
C3C4	45	0,052	84,97	90,10	15,030	75,996	76,00	93,94
C3C4O1	66	0,041	88,45	93,28	11,548	81,479	81,48	95,43
C3C4O1F3	90	0,043	88,83	93,77	11,172	82,019	82,02	95,64
C3C4O1LOCROC	118	0,049	89,08	94,35	10,919	82,218	82,22	95,94
C3C4O1LOCROCX1	**147**	**0,042**	**89,66**	**94,58**	**10,344**	**83,257**	**83,26**	**96,06**
C3C4O1O2	97	0,043	88,73	93,40	11,272	82,134	82,13	95,32
C3C4O1O2F3F4	112	0,046	89,06	93,83	10,944	82,618	82,62	95,49
C3C4O1O2F3F4LOC	114	0,052	88,81	94,02	11,185	81,948	81,95	95,68
C3C4O1O2F3F4LOCROC	110	0,052	88,34	94,07	11,660	80,846	80,85	95,84
C3C4O1O2F3F4LOCROCX1	160	0,047	89,01	94,55	10,993	81,965	81,96	96,05
C3F3O1	99	0,035	89,27	93,10	10,726	83,221	83,22	95,33
C3F3O1LOC	57	0,046	87,76	92,89	12,240	79,716	79,72	95,81
C3F4O2	88	0,044	87,98	91,13	12,020	80,877	80,88	95,08
C3F4O2LOC	105	0,053	88,66	92,30	11,339	82,155	82,16	95,17
C3F4O2LOCROC	75	0,053	88,95	93,46	11,052	82,411	82,41	95,48
C4F3O1	84	0,048	88,41	93,72	11,593	81,136	81,14	95,68
C4F4O2	80	0,060	86,42	91,63	13,575	77,824	77,82	95,03
C4F4O2ROC	98	0,062	87,51	93,44	12,494	79,163	79,16	95,85

BNF: Best Number of Features, ACC: Accuracy, F-me: F-measure, SEN: Sensitivity, SPE: Specificity

Table VII
Algorithm performance in multiclass Sleep Staging with different channels combination

Channels	BNF	CI	AUC	ACC	BER	F-me	SEN	SPE
C3	26	0,025	73,78	85,66	26,220	58,838	57,03	90,66
C3C4	60	0,023	77,30	88,27	22,700	63,333	62,39	92,31
C3C4O1	93	0,024	79,86	89,79	20,136	68,190	66,60	93,19
C3C4O1F3	129	0,033	79,54	89,56	20,464	67,654	66,01	93,11
C3C4O1LOCROC	175	0,017	83,52	91,69	16,477	72,617	72,62	94,44
C3C4O1LOCROCX1	223	0,014	84,10	91,77	15,899	73,778	73,78	94,47
C3C4O1O2	116	0,024	80,32	89,80	19,682	69,971	67,51	93,19
C3C4O1O2F3F4	200	0,033	80,19	89,81	19,807	67,133	67,13	93,28
C3C4O1O2F3F4LOC	230	0,020	82,88	91,28	17,122	73,244	71,64	94,14
C3C4O1O2F3F4LOCROC	264	0,018	83,89	91,78	16,108	73,291	73,29	94,51
C3C4O1O2F3F4LOCROCX1	**326**	**0,015**	**84,67**	**92,04**	**15,329**	**74,738**	**74,74**	**94,64**
C3F3O1	95	0,034	78,59	89,05	21,412	66,583	64,46	92,74
C3F3O1LOC	138	0,025	81,70	90,53	18,297	70,561	69,67	93,66
C3F4O2	100	0,028	78,40	88,63	21,599	64,155	64,16	92,57
C3F4O2LOC	138	0,018	82,64	90,60	17,362	71,381	71,38	93,80
C3F4O2LOCROC	170	0,018	83,40	91,28	16,598	72,540	72,54	94,22
C4F3O1	90	0,032	79,45	89,59	20,553	69,135	65,84	93,09
C4F4O2	98	0,028	78,57	89,01	21,433	66,005	64,49	92,66
C4F4O2ROC	136	0,028	82,44	91,06	17,555	70,902	70,90	94,01

BNF: Best Number of Features, ACC: Accuracy, F-me: F-measure, SEN: Sensitivity, SPE: Specificity

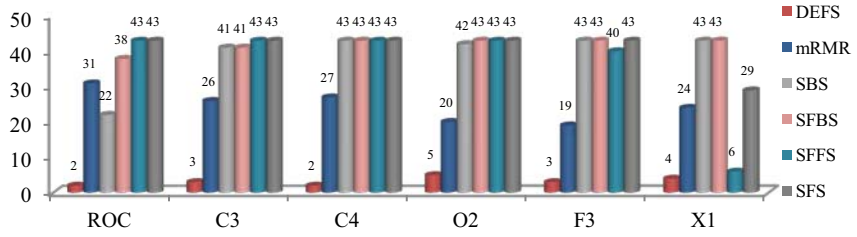

Fig. 6. Number of selected features per channels using different feature selection methods in sleep-awake detection.

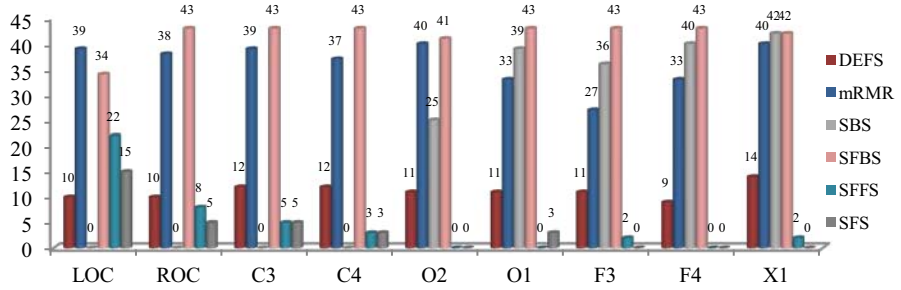

Fig. 7. Number of selected features per channels using different feature selection methods in multiclass sleep staging.

4.3. Channel Selection

Experiments to find the best combination of EEG, EOG and/or EMG channels were performed. Basically, the AASM rules were followed in channel selection. Table VI and Table VII summarize the attained results of different combinations. As highlighted in the tables for sleep-wake detection, the best channels were 3 EEG channels (C3, C4, and O1), 2 EOG channels (ROC and LOC) and 1 EMG channel (X1). On the other hand, the best performance for multiclass sleep staging was achieved using 9 channels: 6 EEG channels (C3, C4, O1, O2, F3 and F4), 2 EOG channels (ROC and LOC) and 1 EMG channel (X1). Furthermore, distributions of the selected features per channel are shown

in Fig. 6 and Fig. 7. These results show the importance of the identified best electrophysiological channels (combinations highlighted in Table VI and Table VII). Moreover, the results confirm the fact that sleep-wake detection is highly dependent on the level of alpha activity in central and occipital channels. The EOG and the EMG channels complemented the information. The EOG should record diverse ocular movements during the wake stage, as the EMG chin channel should record high tonic activity. Furthermore the results of Table VII, confirm the importance of frontal channels in multiclass sleep staging (e.g. in discrimination of REM stage).

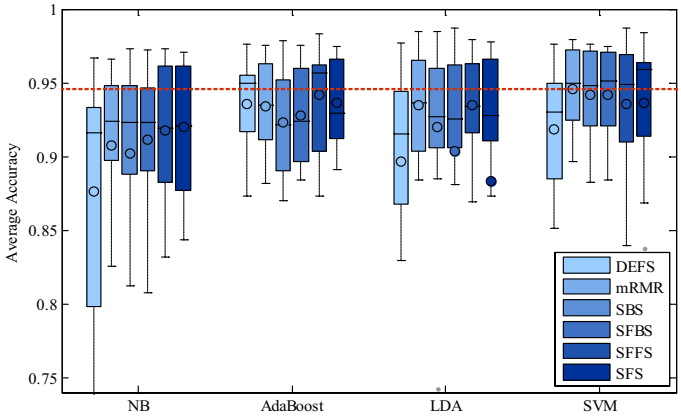

Fig. 8. Accuracy of sleep-wake detection, corresponding to 6 feature selectors (DEFS, mRMR, SBS, SFBS, SFFS and SFS) and 4 classifiers (NB, Adaboost, LDA and SVM).

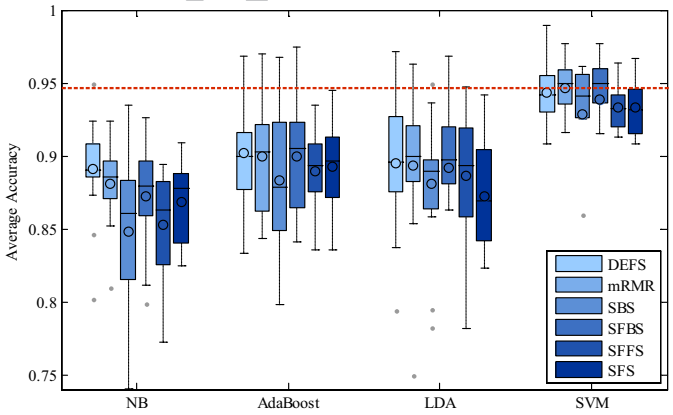

Fig. 9. Accuracy of multiclass sleep staging corresponding to 6 feature selectors (DEFS, mRMR, SBS, SFBS, SFFS and SFS) and 4 classifiers (NB, Adaboost, LDA and SVM).

4.4. Performance Evaluation of the Proposed Scheme

Fig. 8, Fig. 9 compare the performance obtained by the proposed method with different combinations of mentioned feature selector/classifiers detailed in Sections 3.3 and 3.4. In the experiments, six of the best feature selection approaches were used. DEFS, mRMR, and sequential methods (SBS, SFBS, SFFS and SFS). Moreover, four different types of classifiers were considered:

NB, AdaBoost, LDA and SVM classifiers. They are capable of handling large-scale classification problems. The results are expressed in terms of box-whisker plots showing the average, median, the first and third quartile values of the average accuracies. The horizontal lines outside each box identify the upper and lower whiskers, and dot points denote the outliers. It can be observed from the figures that the higher second and third quartiles and the highest average were attained using mRMR-SVM in both sleep-wake detection and multiclass sleep staging. Moreover, as shown in Fig. 9 and Fig. 12, some of other combinations (e.g. DEFS-SVM approach) also perform very close results requiring, however, much less number of features. In multiclass sleep staging the SVM attained the lowest interquartile range and the highest average of accuracies, as shown in Fig. 9. As concerns the sleep-wake detection there is no significant difference between SVM and the other classifiers (see Fig. 8).

4.5. Evaluation of Feature Relevance

To account with the high dimensionality problem and to infer about the most discriminative features, an analysis was performed using our two-step feature -

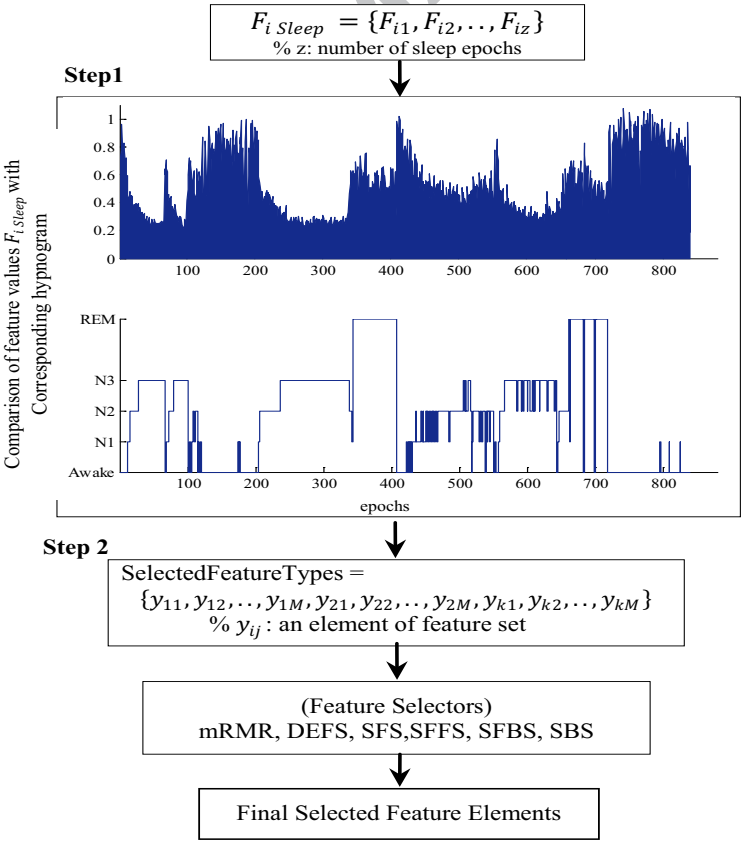

Fig. 10. A sample of the two-step feature selector; step1: histogram of a feature values corresponding to each sleep stages during a whole night hypnogram to select the best feature types; step2: selection of the best feature elements.

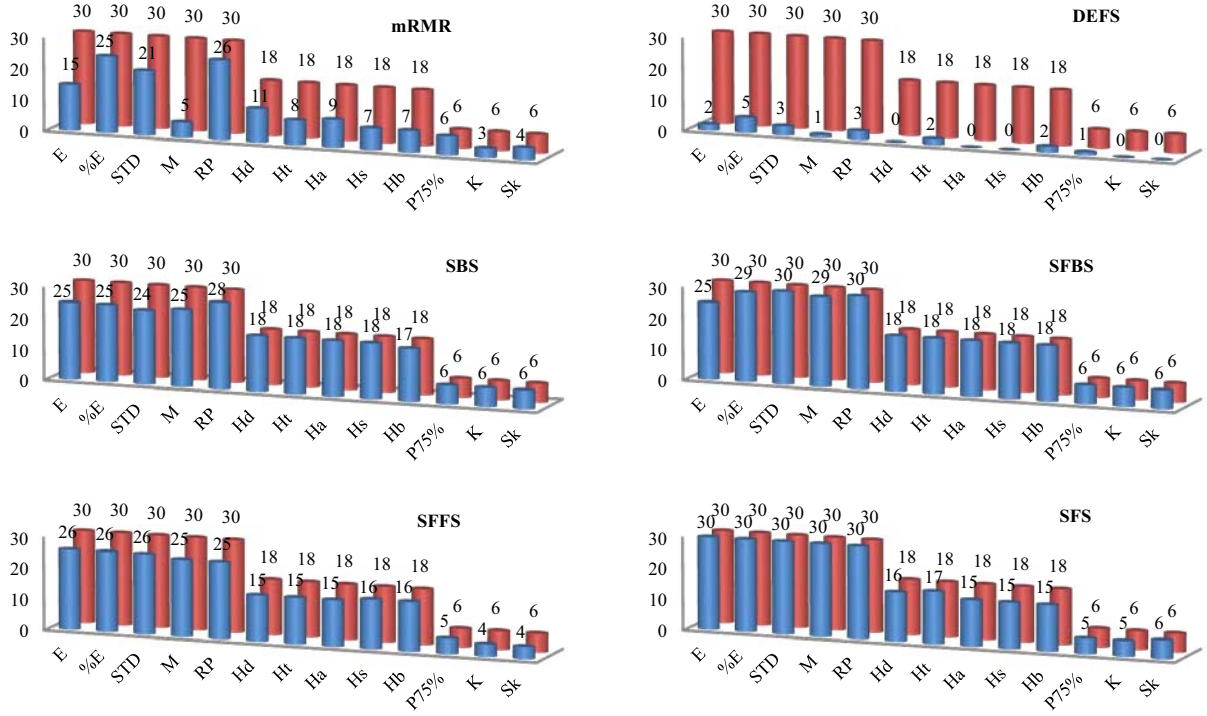

Fig. 11. Selection of the best elements of feature matrix for sleep-awake detection; red: extracted features; blue: selected features. E: energy of sub-bands; %E: percentage of sub-band energy; STD: standard deviation of sub-band energy; M: mean of sub-band energy; RP: relative power; Hd: harmonic-delta; Ht: harmonic-theta; Ha: harmonic alpha; Hs: harmonic-sigma; Hb: harmonic-beta; P75%: percentile 75th; K: kurtosis; Sk: skewness.

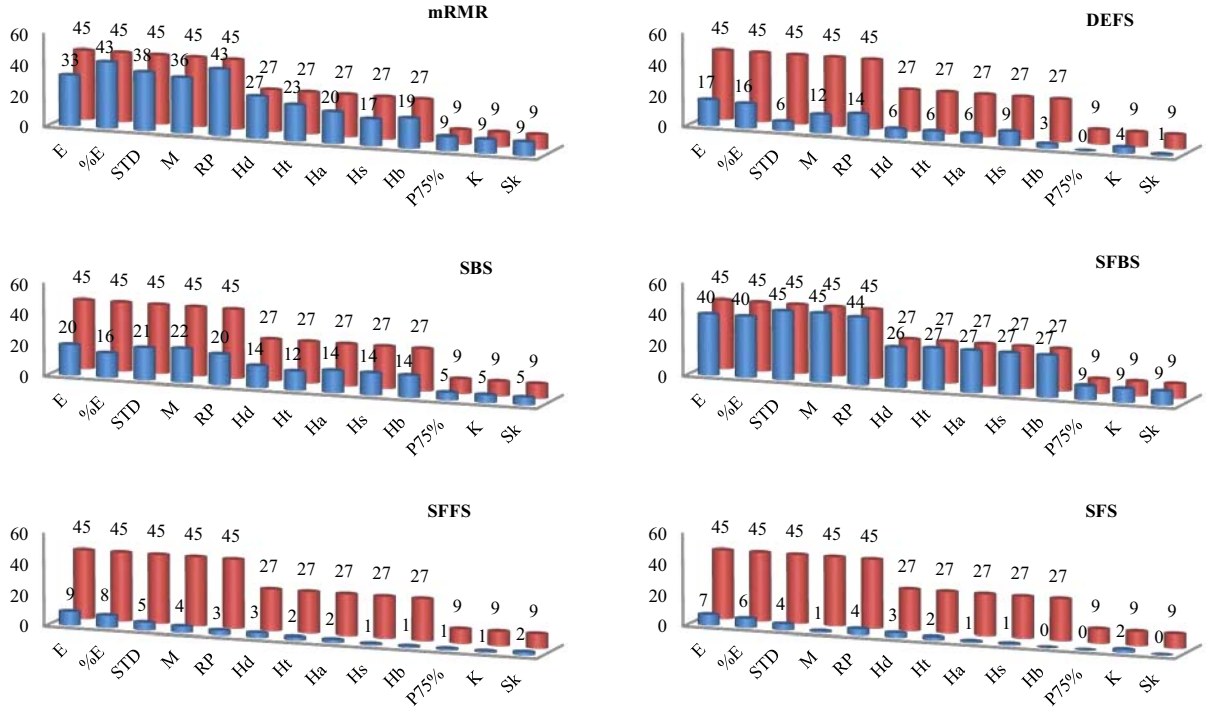

Fig. 12. Selection of the best feature elements for multiclass sleep staging; red: extracted features; blue: selected features. E: energy of sub-bands; %E: percentage of sub-band energy; STD: standard deviation of sub-band energy; M: mean of sub-band energy; RP: relative power; Hd: harmonic-delta; Ht: harmonic-theta; Ha: harmonic alpha; Hs: harmonic-sigma; Hb: harmonic-beta; P75%: percentile 75th; K: kurtosis; Sk: skewness.

selection approach (See Fig. 10). Firstly, as detailed in Algorithm 1, some of the extracted feature types were selected manually. By analyzing on the histogram of features distribution of whole night sleep and corresponding hypnogram, we inferred the following types of features as being the most discriminative: MODWT based features (energy, percentage of energy, mean and standard deviation of sub-bands), and relative-power, harmonic of theta, sigma, beta and alpha frequency ranges, percentile 75%, kurtosis and skewness. The second step is carried out with the purpose of selecting the final feature elements, i.e., for each feature type the final elements are selected. Therefore, resulted features of the first step, are fed into the feature selectors mentioned in Section 3.3. As illustrated in Fig. 11 and Fig. 12 relative-power and percentage-of-energy are the most discriminative features for both sleep-wake detection and multiclass sleep staging. Moreover, it can be inferred from indicated figures that all features which selected in the first step are important and useful for the classification phase.

4.6. Analysis by gender

In other to evaluate the performance of the proposed method by gender, two experiments were performed: 1) the proposed ASSC was trained with data of subjects of both genders; 2) the ASSC system was trained and tested separately per each type of gender. Fig. 13 and Fig. 14 provide the accuracies, F-measures, and specificities, obtained with the ISRUC-Sleep dataset (see Table VIII), comprising 40 subjects, 14 female and 26 male subjects. The ASSC method achieved a better performance for both applications when trained/tested separately by gender. Actually, in both applications we had a lower interquartile range on the accuracy, F-measure and specificity of when the ASSC system was trained and tested separately by gender. Moreover, no significant differences were found in accuracy, F-measure, or specificity between female and male subjects.

4.7. Global performance of the proposed scheme

Table VIII and Table IX summarize the details of the overall performance of the proposed ASSC method using the ISRUC-Sleep dataset, on sleep-wake detection and multiclass sleep staging applications. The three last columns of Table VIII provide sensitivity, specificity and accuracy values for each of the 40 subjects, in the case of sleep-wake detection. For sleep-wake detection the following average sensitivity (93.74), specificity (82.49) and accuracy (88.87)

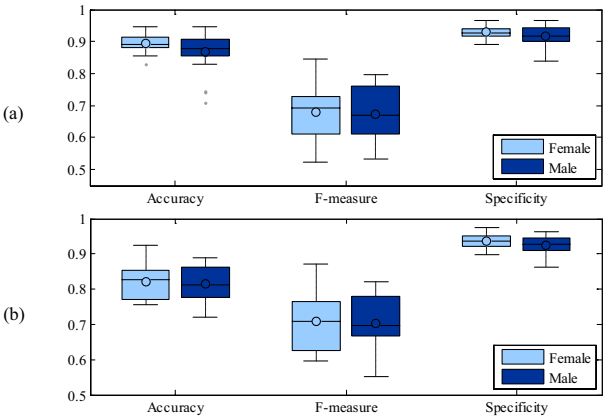

Fig. 13. Accuracy F-measure and specificity of sleep-wake detection correspond to (a) training and test with the same genders (b) training with the both genders.

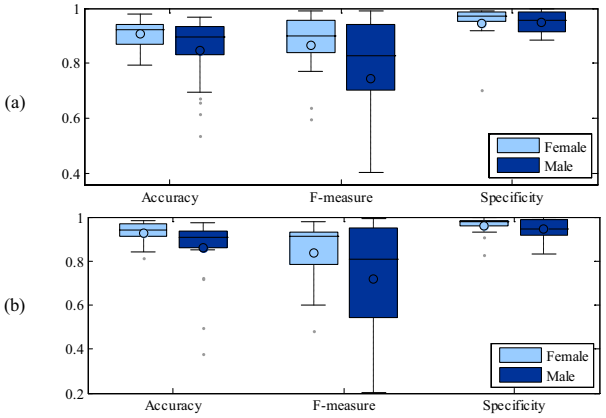

Fig. 14. Accuracy, F-measure and specificity of multiclass sleep staging correspond to (a) training and test with the same genders (b) training with the both genders.

were achieved. The lowest sensitivity values were obtained for the subjects with longer periods of wake stage during a whole night recording (approximately 8 hours of data collection) (e.g. Subjects 6 and13 in Table VIII). On the other hand, as it can be seen from data of Table IX , in multiclass sleep staging, the best discrimination was achieved for awake (average sensitivity of 84.12 and average specificity of 93.07) N3 (average sensitivity of 78.51 and average specificity of 95.75) and REM stages (average sensitivity of 79.30 and average specificity of 94.69). However, the lowest sensitivities reside in the detection of stages N1 (41.71), while the average specificities attained for N1 (92.11) are close to the other sleep stages. Indeed, attending to accuracy values, it can be said that the best results were achieved regarding awake (88.59) and N3 (87.13), followed by REM (86.99), N2 (79.06) and N1 (66.91). The results indicate the performance of the proposed method.

Table VIII

Details of Our ISRUC-Sleep dataset and Performance results to the sleep-wake detection.

Subject	Age	Sex	Sleep Apnea event	Other problems	nº of epoch	% W	% N1	% N2	% N3	% REM	Sens	Spec	Acc
1	64	M	yes	overweight	880	30.00	8.30	22.05	26.25	13.41	90.70	99.32	95.01
2	52	M	yes	overweight	964	25.41	11.93	35.79	16.29	10.58	83.68	99.14	91.41
3	38	M	no	parasomnia	943	14.00	17.50	26.09	18.35	24.07	81.54	95.91	88.73
4	27	M	yes	-	963	2.91	6.75	44.24	22.22	23.88	82.14	97.35	89.75
5	58	F	yes	insomnia	875	33.83	12.34	30.29	18.74	4.80	95.49	98.62	97.05
6	22	M	no	PLMS and insomnia	897	80.49	1.78	6.69	11.04	0.00	31.94	99.43	65.68
7	70	M	yes	ICC	933	14.36	18.33	21.97	25.08	20.26	53.73	95.58	89.37
8	76	M	yes	PLMS	904	24.45	13.94	31.08	23.67	6.86	88.08	94.21	93.25
9	61	M	yes	overweight	969	15.48	17.85	35.19	16.41	15.07	80.17	92.79	86.48
10	53	F	yes	PLMS, AVC overweight	842	38.00	10.69	36.58	11.40	3.33	91.03	95.21	93.12
11	57	F	yes	narcolepsy	982	24.44	15.89	45.42	9.16	5.09	99.15	70.15	84.65
12	79	M	yes	-	850	19.65	9.29	17.53	39.29	14.24	99.15	56.21	66.81
13	65	M	yes	leukemia and HTA	882	73.70	12.59	4.42	7.26	1.93	24.65	99.41	86.46
14	66	M	yes	ataxia	906	50.77	12.36	19.98	11.48	5.41	97.38	89.00	93.19
15	52	M	yes	depression	786	24.55	10.81	20.10	23.03	21.50	90.53	93.99	92.26
16	50	M	yes	-	883	17.78	17.44	37.26	13.59	13.93	94.16	88.69	91.42
17	79	M	yes	-	851	44.18	18.68	24.91	6.46	5.76	89.37	90.09	89.73
18	38	M	yes	PLMS	999	13.61	10.81	43.84	15.42	16.32	96.30	97.24	96.77
19	59	F	yes	HTA, obesity and diabetes	828	40.82	18.12	23.07	8.21	9.78	90.40	91.79	91.10
20	59	M	yes	HTA and epilepsy	692	24.57	9.83	15.32	38.01	12.28	96.60	92.21	94.41
21	72	F	yes	obesity and diabetes	1054	35.77	13.00	17.08	13.38	20.78	63.93	97.22	80.57
22	85	M	yes	Alzheimer	849	36.04	7.89	35.45	12.60	8.01	71.03	95.46	83.25
23	50	F	yes	PLMS	892	29.82	12.78	34.98	7.29	15.13	88.30	96.65	92.48
24	65	M	yes	HTA	830	24.10	10.24	29.40	16.14	20.12	73.53	99.52	86.53
25	29	F	no	-	921	14.77	6.62	31.92	14.55	32.14	59.56	98.94	79.25
26	69	M	yes	bronchitis	1062	28.53	20.24	19.59	15.54	16.10	93.07	88.34	90.70
27	26	F	yes	PLMS and obesity	914	32.71	6.24	23.74	26.04	11.27	97.00	98.81	97.90
28	62	F	yes	-	882	6.93	11.02	19.66	24.66	37.73	77.05	96.97	87.01
29	42	F	yes	obesity	912	26.43	22.26	28.40	19.08	3.84	86.46	97.40	91.93
30	51	M	yes	HTA	882	30.76	14.07	39.05	8.06	8.06	96.31	91.91	94.11
31	29	M	yes	-	877	9.64	15.60	47.26	14.05	13.45	78.57	91.35	84.96
32	65	M	yes	-	1030	5.71	22.50	39.88	15.48	16.43	89.80	97.53	93.66
33	32	F	yes	obesity and PLMS	920	44.52	6.43	31.79	10.71	6.55	94.10	98.93	96.52
34	43	F	yes	PLMS and HTA	871	8.21	17.74	38.45	20.83	14.76	89.86	97.28	93.57
35	59	M	yes	HTA	888	40.95	11.19	22.86	19.64	5.36	40.45	98.65	69.55
36	36	F	yes	affective disorder	987	33.33	10.36	22.62	14.88	18.81	83.75	95.70	89.72
37	52	M	yes	-	806	24.49	15.87	29.48	13.49	16.67	70.37	98.30	84.33
38	37	M	yes	dyslipidemia	932	11.56	23.70	39.68	8.62	16.44	99.08	90.67	94.88
39	66	M	yes	polycythemia	900	37.83	13.14	20.69	18.86	9.49	94.93	91.59	93.26
40	62	F	yes	PLMS	875	63.31	2.63	10.51	22.63	0.91	96.57	91.75	94.16

Sens: Sensitivity, Spec: Specificity, Acc: Accuracy

Table IX
Performance results to the multiclass sleep staging.

	Wake			N1			N2			N3			REM			Average		
Subjects	**Sens**	**Spec**	**Acc**	**Sens**	**Spec**	**Acc**	**Sens**	**Spec**	**Acc**	**Sens**	**Spec**	**Acc**	**Sens**	**Spec**	**Acc**	**Sens**	**Spec**	**Acc**
1	74.01	98.14	96.36	27.14	98.08	62.61	86.71	89.36	88.04	89.18	96.45	92.81	92.37	99.73	96.05	**77.99**	**96.35**	**87.17**
2	88.70	97.84	93.27	55.34	93.50	74.42	76.74	89.15	82.95	80.89	94.85	87.87	92.31	97.98	95.15	**78.80**	**94.67**	**86.73**
3	86.92	92.46	89.69	54.94	93.21	74.07	87.40	90.40	88.90	77.46	99.86	88.66	82.67	97.19	89.93	**77.88**	**94.63**	**86.25**
4	89.29	96.35	92.82	42.86	97.13	69.99	80.19	90.18	85.18	92.52	91.93	92.23	79.90	98.08	88.99	**76.95**	**94.73**	**85.84**
5	99.25	96.37	97.81	50.00	98.37	74.19	83.40	92.59	87.99	95.12	99.41	97.27	90.48	96.01	93.25	**83.65**	**96.55**	**90.10**
6	40.90	99.43	70.16	93.75	98.47	96.11	36.67	100.00	68.33	100.00	95.57	97.79	100.00	53.86	76.93	**74.26**	**89.47**	**81.87**
7	75.37	96.88	86.13	6.43	97.81	52.12	8.78	95.42	52.10	94.87	86.40	90.63	94.97	68.15	81.56	**56.09**	**88.93**	**72.51**
8	63.73	93.83	78.78	16.94	88.00	52.47	29.54	92.92	61.23	84.11	93.18	88.65	82.26	75.74	79.00	**55.31**	**88.73**	**72.02**
9	78.51	88.88	83.69	31.40	95.57	63.48	85.04	82.61	83.83	61.01	99.23	80.12	90.41	95.46	92.94	**69.27**	**92.35**	**80.81**
10	84.14	93.87	89.00	36.67	96.81	66.74	90.26	78.57	84.42	1.04	99.44	50.24	85.71	91.71	88.71	**59.56**	**92.08**	**75.82**
11	100.00	59.83	79.92	1.92	97.86	49.89	65.25	92.09	78.67	60.00	99.88	79.94	80.00	99.68	89.84	**61.43**	**89.87**	**75.65**
12	33.80	99.56	66.68	63.51	89.95	76.73	65.77	90.61	78.19	94.01	90.33	92.17	86.78	97.14	91.96	**68.78**	**93.52**	**81.15**
13	29.31	100.00	64.65	32.43	50.20	41.32	89.74	89.91	89.83	100.00	99.49	99.75	94.12	92.34	93.23	**69.12**	**86.39**	**77.75**
14	98.03	87.80	92.92	29.46	97.25	63.36	79.55	94.43	86.99	99.04	98.32	98.68	92.31	99.65	95.98	**79.68**	**95.49**	**87.58**
15	91.05	93.64	92.35	32.10	95.26	63.68	79.87	80.23	80.05	80.11	96.70	88.40	50.00	98.68	74.34	**66.63**	**92.90**	**79.76**
16	96.35	85.61	90.98	35.29	86.14	60.72	56.23	92.75	74.49	98.33	92.09	95.21	58.77	99.86	79.32	**69.00**	**91.29**	**80.14**
17	90.74	89.87	90.30	42.28	86.61	64.44	62.69	86.45	74.57	80.00	98.04	89.02	32.65	99.48	66.07	**61.67**	**92.09**	**76.88**
18	94.81	97.36	96.09	35.85	98.96	67.40	92.70	86.02	89.36	77.92	98.53	88.22	96.32	97.02	96.67	**79.52**	**95.58**	**87.55**
19	91.02	86.11	88.56	12.67	98.92	55.79	78.53	83.86	81.19	92.65	95.62	94.13	89.39	98.63	94.01	**72.85**	**92.63**	**82.74**
20	98.11	89.16	93.64	41.24	92.35	66.79	57.62	82.83	70.23	58.58	100.00	79.29	80.61	99.15	89.88	**67.23**	**92.70**	**79.96**
21	68.44	96.14	82.29	11.76	89.30	50.53	43.02	86.85	64.94	100.00	88.83	94.41	89.50	96.65	93.07	**62.54**	**91.55**	**77.05**
22	92.41	81.10	86.76	25.00	93.32	59.16	56.04	88.48	72.26	63.55	94.80	79.18	47.06	97.07	72.06	**56.81**	**90.95**	**73.88**
23	95.85	91.29	93.57	26.36	96.28	61.32	82.76	83.22	82.99	70.77	94.35	82.56	54.55	100.00	77.27	**66.06**	**93.03**	**79.54**
24	81.76	98.10	89.93	55.29	88.53	71.91	73.36	94.60	83.98	98.51	96.25	97.38	76.65	95.89	86.27	**77.11**	**94.67**	**85.89**
25	60.29	97.88	79.09	27.27	97.37	62.32	40.07	91.89	65.98	97.01	68.69	82.85	77.42	98.53	87.97	**60.41**	**90.87**	**75.64**
26	95.05	91.36	93.20	77.46	84.25	80.86	53.37	79.37	66.37	6.67	100.00	53.33	57.34	98.54	77.94	**57.98**	**90.70**	**74.34**
27	94.67	98.81	96.74	19.64	99.16	59.40	94.18	83.40	88.79	56.90	100.00	78.45	95.19	93.62	94.41	**72.12**	**95.00**	**83.56**
28	93.44	98.36	95.90	11.70	99.21	55.46	35.06	92.33	63.69	98.16	84.72	91.44	96.41	91.39	93.90	**66.95**	**93.20**	**80.08**
29	94.32	90.66	92.49	49.74	92.19	70.96	76.36	91.51	83.93	95.40	98.02	96.71	66.67	99.30	82.98	**76.50**	**94.33**	**85.42**
30	95.94	91.74	93.84	49.59	92.07	70.83	58.59	97.72	78.15	98.61	91.41	95.01	96.77	96.84	96.80	**79.90**	**93.95**	**86.93**
31	82.14	89.52	85.83	54.96	83.94	69.45	62.09	74.22	68.16	27.12	99.86	63.49	75.22	96.05	85.64	**60.31**	**88.72**	**74.51**
32	93.88	96.35	95.11	65.07	85.21	75.14	74.45	77.57	76.01	77.69	98.82	88.26	57.25	99.52	78.39	**73.67**	**91.50**	**82.58**
33	96.70	97.64	97.17	66.67	94.74	80.70	82.02	96.63	89.33	92.22	98.50	95.36	98.18	100.00	99.09	**87.16**	**97.50**	**92.33**
34	89.86	96.11	92.98	70.67	87.70	79.18	75.85	88.61	82.23	82.29	99.40	90.84	79.03	98.88	88.96	**79.54**	**94.14**	**86.84**
35	73.89	96.40	85.14	78.41	91.04	84.73	71.10	97.44	84.27	94.93	96.94	95.93	60.00	90.74	75.37	**75.66**	**94.51**	**85.09**
36	85.87	95.10	90.48	47.57	87.59	67.58	70.81	91.18	80.99	78.63	99.64	89.13	86.58	99.17	92.88	**73.89**	**94.54**	**84.21**
37	60.32	98.64	79.48	51.67	93.75	72.71	84.80	86.69	85.75	75.51	99.26	87.39	74.79	84.63	79.71	**69.42**	**92.59**	**81.01**
38	99.08	91.80	95.44	56.81	87.81	72.31	77.72	83.79	80.75	46.05	99.64	72.84	68.97	97.49	83.23	**69.72**	**92.11**	**80.92**
39	94.93	90.65	92.79	63.24	94.55	78.89	80.11	97.81	88.96	96.92	97.16	97.04	74.70	99.62	87.16	**81.98**	**95.96**	**88.97**
40	91.16	92.10	91.63	17.39	95.99	56.69	84.78	92.30	88.54	66.67	98.52	82.59	87.50	98.21	92.85	**69.50**	**95.42**	**82.46**
Total	**84.12**	**93.07**	**88.59**	**41.71**	**92.11**	**66.91**	**69.23**	**88.90**	**79.06**	**78.51**	**95.75**	**87.13**	**79.30**	**94.69**	**86.99**	**70.57**	**92.90**	**81.74**

Sens: Sensitivity, Spec: Specificity, Acc: Accuracy

5. Discussion and Conclusion

To discriminate the sleep stages based on AASM standard, a subject-independent ASSC method was here proposed for sleep-wake detection and for multiclass sleep staging (awake, NREM (N1, N2, N3) sleep and REM sleep). The method employs the advantages of extracted features from multi-channels EEG, EOG and EMG signals according to temporal, frequency and time-frequency domains. Applying the MODWT, which omitted subsampling in the filtering process, provided the shift invariance characteristic to our method which is one of the most important properties in analysis of PSG signals. To reduce the effect of extreme values in the feature vectors the extracted feature

set was transformed and normalized, which improved the overall performance (Fig. 3). Moreover, by using the two-step feature selector it was inferred that, relative-power and percentage-of-energy are the most discriminative features for both sleep-wake detection and multiclass sleep staging (Fig. 11and Fig. 12). The proposed method perform the best performance by combining 6 channels (C3, C4, O1, ROC, LOC and X1) for sleep-wake detection, and 9 channels (C3, C4, O1, O2, F3, F4, ROC, LOC, X1) for multiclass sleep staging (Table VI and Table VII). The experimental study was performed using the ISRUC-Sleep dataset which is a rich dataset composed by PSG signals from 40 subjects with different characteristics (i.e. young/old, male/female, non-apnea/with apnea event and other sleep problems). The overall accuracy of the proposed method applied to PSG signals from 40 subjects reached 88.87 and 81.74 for sleep-wake detection and multiclass sleep staging, respectively (Table VIII and Table IX). As concerns REM stage, remarkable results have been achieved by our ASSC method (accuracy of 86.99, specificity 94.69 and sensitivity 79.30). In fact it was verified that employing EOG and EMG channels, improved the REM stage discrimination in comparison with only using EEG channels. Indeed, in REM stage EOG signals capture the high ocular activity (rapid eye movements) and the EMG signal captures the low level of (chin) muscle tone (the opposite of awake stage (Table I)). On the other hand, the worst accuracies occurred for N1, N2 stages. Indeed, recognition of N1 is one of the main challenges of sleep staging. There is a lack of discriminative features that characterize N1 stage clearly from the other stages. This has been observed previously by many authors (e.g., Anderer *et al.* [40]). This could be due to N1 being a transition phase between wakefulness and different sleep stages as discussed in [41]. In fact, the neurophysiologic signals of N1 and N2 stages present similarities between themselves and a mix of patterns with similarities to awake, N3 and REM stages (Table I); e.g., the N1 epochs can present alpha activity (typical of awake stage) and can present theta activity (typical of N2 sleep stage). Moreover, most of the times the N2 sleep stage automatically is misclassified as N1 or N3. Furthermore, concerning pattern similarities between N2 with N3, critical cases are the transition epochs: epochs with a relevant percentage of slow waves but not enough to be classified as N3 sleep stage (Table I). Finally, the worst cases of performance in multiclass sleep staging were mainly related to the older subjects with the high percentages of epochs in N1 and N2 sleep stages (e.g. subjects 7, 8 and 22 in Table IX).

References

[1] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, and F. Chapotot, "Feature selection for sleep/wake stages classification using data driven methods," *Biomedical Signal Processing and Control,* vol. 2, pp. 171-179, Jul 2007.

[2] A.Ambrogetti, M. J. Hensley, and L.G.Olsen, *Sleep Disorders: A Clinical Textbook*. London: Quay Books, 2006.

[3] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," ed. Bethesda, MD, U.S. National Institute of Neurological Diseases and Blindness, Neurol. Inform. Netw., 1968.

[4] C. Iber, S. Ancoli-Israel, A. Chesson, and S. Quan, *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*, 2007.

[5] M. H. Silber, S. Ancoli-Israel, M. H. Bonnet, S. Chokroverty, M. M. Grigg-Damberger, M. Hirshkowitz, S. Kapen, S. A. Keenan, M. H. Kryger, T. Penzel, M. R. Pressman, and C. Iber, "The visual scoring of sleep in adults," *J Clin Sleep Med,* vol. 3, pp. 121-31, Mar 15 2007.

[6] H. G. Jo, J. Y. Park, C. K. Lee, S. K. An, and S. K. Yoo, "Genetic fuzzy classifier for sleep stage identification," *Computers in Biology and Medicine,* vol. 40, pp. 629-634, Jul 2010.

[7] W. C. Tang, S. W. Lu, C. M. Tsai, C. Y. Kao, and H. H. Lee, "Harmonic Parameters with HHT and Wavelet Transform for Automatic Sleep Stages Scoring," *Proceedings of World Academy of Science, Engineering and Technology, Vol 22,* vol. 22, pp. 414-417, 2007.

[8] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," *Comuter Methods and programs in biomedicine,* 2011.

[9] S. Gunes, K. Polat, and S. Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting," *Expert Systems with Applications,* vol. 37, pp. 7922-7928, Dec 2010.

[10] S. Khalighi, T. Sousa, D. Oliveira, G. Pires, and U. Nunes, "Efficient Feature Selection for Sleep Staging Based on Maximal Overlap Discrete Wavelet Transform and SVM," *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* pp. 3306-3309, 2011.

[11] S. Khalighi, T. Sousa, and U. Nunes, "Adaptive automatic sleep stage classification under covariate shift," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 2259-62.

[12] M. Ronzhina, O. Janousek, J. Kolarova, M. Novakova, P. Honzik, and I. Provaznik, "Sleep scoring using artificial neural networks," *Sleep Medicine Reviews,* vol. 16, pp. 251-263, Jun 2012.

[13] L. Fraiwan, K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, and H. Dickhaus, "Classification of Sleep Stages Using Multi-wavelet Time Frequency Entropy and LDA," *Methods of Information in Medicine,* vol. 49, pp. 230-237, 2010.

[14] D. Álvarez-Estévez, J. M. Fernández-Pastoriza, E. Hernández-Pereira, and V. Moret-Bonillo, "A method for the automatic analysis of the sleep macrostructure in continuum," *Expert Systems with Applications,* vol. 40, pp. 1796–1803, 2013 2013.

[15] V. C. F. Helland, A. Gapelyuk, A. Suhrbier, M. Riedl, T. Penzel, J. Kurths, and N. Wessel, "Investigation of an Automatic Sleep Stage Classification by Means of Multiscorer Hypnogram," *Methods of Information in Medicine,* vol. 49, pp. 467-472, 2010.

[16] M. E. Tagluk, N. Sezgin, and M. Akin, "Estimation of Sleep Stages by an Artificial Neural Network Employing EEG, EMG and EOG," *Journal of Medical Systems,* vol. 34, pp. 717-725, Aug 2010.

[17] F. Chapotot and G. Becq, "Automated sleep-wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules," *International Journal of Adaptive Control and Signal Processing,* vol. 24, pp. 409-423, May 2010.

[18] B. Kemp and J. Olivan, "European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data," *Clinical Neurophysiology,* vol. 114, pp. 1755-1761, Sep 2003.

[19] H. Jasper, "Appendix to report to committee on clinical examination in EEG: the ten–twenty electrode system of the international federation," *Electroencephalography and Clinical Neurophysiology,* pp. 371-375, 1958.

[20] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *Journal of Neuroscience Methods,* vol. 123, pp. 69-87, Feb 15 2003.

[21] B. D.Percival and T. A.Walden, *Wavelet Methods for Time Series Analysis*: Cambridge University Press, 2000.

[22] S. Ahmad, "Temporal Pattern Identification and Summarization Method for Complex Time Serial Data," PhD dessertation, Information Extraction and Multimedia Group Dep. of Comp. School of Elec. and Phy. Sci., Uni.of Surrey Guildford, Surrey GU2 7XH, UK., 2007.

[23] I. Omerhodzic, S. Avdakovic, A. Nuhanovic, and K. Dizdarevic, "Energy Distribution of EEG Signals: EEG Signal Wavelet-Neural Network Classifier," *International Journal of Biological and Life Sciences,* vol. 6, 2010.

[24] A. S. Yilmaz, A. Alkan, and M. H. Asyali, "Applications of parametric spectral estimation methods on detection of power system harmonics," *Electric Power Systems Research,* vol. 78, pp. 683-693, Apr 2008.

[25] E. Arbabi and M. Shamsollahi, "Comparation between different types of features used for classification in BCI," in *12th International Conference on Biomedical Engineering, ICBME'05*, 2005.

[26] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal,* vol. 27, pp. 623-656, 1948.

[27] A. Renyi, "On measures of entropy and information," in *Fourth Berkeley Symp. Math. Stat. Prob.*, Berkely, 1960, p. 547.

[28] T. Maszczyk and W. Duch, "Comparison of Shannon, Renyi and Tsallis entropy used in decision trees," *Artificial Intelligence and Soft Computing - ICAISC 2008, Proceedings,* vol. 5097, pp. 643-651, 2008.

[29] C. Tsallis, R. S. Mendes, and A. R. Plastino, "The role of constraints within generalized nonextensive statistics," *Physica A,* vol. 261, pp. 534-554, Dec 15 1998.

[30] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain,* vol. 130, pp. 314-333, Feb 2007.

[31] R. Agarwal and J. Gotman, "Computer-assisted sleep staging," *IEEE Transactions on Biomedical Engineering,* vol. 48, pp. 1412-1423, Dec 2001.

[32] K. Ansari-Asl, G. Chanel, and T. Pun, "A Channel Selection Method for EEG Classification in Emotion Assessment based on

Synchronization likelihood," presented at the 15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, 2007.

[33] G. Becq, S. Charbonnier, F. Chapotot, A. Buguet, L. Bourdon, and P. Baconnier, "Comparison Between Five Classifiers for Automatic Scoring of Human Sleep Recordings," in *Studies in Computational Intelligence (SCI)* ed: Springer, 2005, pp. 113–127.

[34] H. C. Peng, F. H. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 1226-1238, Aug 2005.

[35] A. W. Whitney, "A Direct Method of Nonparametric Measurement Selection," *IEEE Trans. on Comput.,* vol. 20, pp. 1100 – 1103, 1971.

[36] P. Pudil, J. Novoviˇcová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters,* vol. 15, pp. 1119 – 1125, 1994.

[37] R. N. Khushaba, A. Al-Ani, and A. Al-Jumaily, "Feature subset selection using differential evolution and a statistical repair mechanism," *Expert Systems with Applications,* vol. 38, pp. 11515-11526, Sep 2011.

[38] J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery,* vol. 2, 1998.

[39] C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines,," *ACM Transactions on Intelligent Systems and Technology,* vol. 2, p. 27, 2011.

[40] P. Anderer, G. Gruber, S. Parapatics, M. Woertz, T. Miazhynskaia, G. Klosch, B. Saletu, J. Zeitlhofer, M. J. Barbanoj, H. Danker-Hopfe, S. L. Himanen, B. Kemp, T. Penzel, M. Grozinger, D. Kunz, P. Rappelsberger, A. Schlogl, and G. Dorffner, "An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: Validation study of the Somnolyzer 24 x 7 utilizing the Siesta database," *Neuropsychobiology,* vol. 51, pp. 115-133, 2005.

[41] S. L. Himanen and J. Hasan, "Limitations of Rechtschaffen and Kales," *Sleep Medicine Reviews,* vol. 4, pp. 149-167, Apr 2000.

Research Highlights

- A subject-independent automatic sleep staging method with application in sleep-wake detection and in multiclass sleep staging.
- An extensive dataset with 40 polysomnographic (PSG) recording.
- A time-frequency based feature extraction method using maximum overlap discrete wavelet transform (MODWT).
- A two-step feature selector to find the most discriminative features.
- The best combinations of the PSG channels in sleep-wake detection and in multiclass sleep staging.