A Comparative Review on Sleep Stage Classification Methods in Patients and healthy Individuals

Reza Boostani, Foroozan Karimzadeh, Mohammad Torabi-Nami

A Comparative Review on Sleep Stage Classification Methods in Patients and healthy Individuals

Reza Boostani[a], Foroozan Karimzadeh[b*], Mohammad Torabi-Nami[c]

[a,b] *Department of Computer Science and Information technology, School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran.*
[c] *Department of Neuroscience, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran.*
boostani@shirazu.ac.ir , torabinami@sums.ac.ir
[*] *Corresponding author: Tel: +987136474605, E-mail: f.karimzadeh@cse.shirazu.ac.ir*

Abstract

Background and objective: Proper scoring of sleep stages can give clinical information on diagnosing patients with sleep disorders. Since traditional visual scoring of the entire sleep is highly time-consuming and dependent to experts' experience, automatic schemes based on electroencephalogram (EEG) analysis are broadly developed to solve these problems. This review presents an overview on the most suitable methods in terms of preprocessing, feature extraction, feature selection and classifier adopted to precisely discriminate the sleep stages.

Methods: This study round up a wide range of research findings concerning the application of the sleep stage classification. The fundamental qualitative methods along with the state-of-the-art quantitative techniques for sleep stage scoring are comprehensively introduced. Moreover, according to the results of the investigated studies, five research papers are chosen and practically implemented on a well-known public available sleep EEG dataset. They are applied to single-channel EEG of 20 subjects containing equal number of healthy and patient individuals. Feature extraction and classification schemes are assessed in terms of accuracy and robustness against noise. Furthermore, an additional implementation phase is added to this research in which all combinations of the implemented features and classifiers are considered to find the best combination for sleep analysis.

Results: According to our achieved results on both groups, entropy of wavelet coefficients along with random forest classifier are chosen as the best feature and classifier, respectively. The mentioned feature and classifier provide 88.66% accuracy on healthy subjects and 66.96% on patient group. Therefore, it can be claimed that continuous wavelet features and random forest provide the best result on this dataset.

Conclusions: In this paper, the road map of EEG-base sleep stage scoring methods is clearly sketched. Implementing the state-of-the-art methods and even their combination on both healthy and patient datasets indicates that although the accuracy on healthy subjects are remarkable, the results for the main community (patient group) by the quantitative methods are not promising yet. The reasons rise from adopting non-matched sleep EEG features from other signal processing fields such as communication. As a conclusion, developing sleep pattern-related features deem necessary to enhance the performance of this process.

Keywords: Sleep stage classification, wavelet transform, random forest classifier, entropy.

1. Introduction

Sleep covers almost one third of human lifespan. Due to the direct relationship among sleep quality and humans' physical and mental performance, sufficient night sleep is crucial. As a result of machinery and stressful life, sleep disturbance is increasing in modern societies. In addition, research findings suggest that several psychological and neurological disorders can deteriorate normal sleep patterns [1]. According to the international classification of sleep disorders (ICSD-II) criteria [2], eighty four different sleep disorders are defined. Sleep disorders not only cause a reduction in physical performance during the day, but also leave negative effects on cognitive functions such as attention, learning and memory, in long-term [3]. For instance, beside the significant side effects of obstructive sleep apnea syndrome (OSAS) including the increased risk of cardiovascular diseases; neurocognitive declines and excessive daytime are considered as potential consequences [3].

To achieve the right diagnosis and treatment based on the various biological records, accurate sleep scoring is deemed to be a crucial part of the process. Up to now, the conventional visual scoring method is still the most acceptable approach, though it involves visual data interpretation of different signals [4]. Qualitative scoring, however, subject to some pitfalls including experts' experience which might result in different scoring results by different experts [5, 6]. In an optimistic view, the agreement between the obtained results by two experts, in average, is 83 ± 3 % [7] which is not convincing. In addition, visual inspection is a time-consuming process for a whole night EEG labeling. Therefore, automatic scoring is deemed to be an efficient approach [8, 9].

Several research teams have recently proposed various methods to automate the process of sleep classification (sleep scor-

ing). Several signal processing techniques along with machine learning algorithms are adopted to obtain useful information from biological signals [10]. Such methods are divided into two categories, i.e. multi-channel and single-channel processing. In the former approach, the combination of various biological signals such as multi-channel EEG signals, electromyogram (EMG) [11] and electrooculogram (EOG) are utilized to extract informative features [12, 13, 14, 15, 16, 17, 18]. While the use of multi-channel signals leads to the higher performance [19], it imposes a considerable cost to patients, especially in home sleep testing [5]. Moreover, excessive number of wire connections during the recording process might per se result in sleep disturbance [20].

On the other hand, single-channel EEG based analysis is a cheap way of automatic sleep scoring. EEG contains valuable and interpretable information resembling the brain activities which is not only used in extensive research contexts pertaining to the brain, but also to diagnose and consequently the treatment of neurological disorders [21]. Sleep neurology is a progressively-evolving sub-specialty field in which the sleep EEG signals are utilized to study the function of the brain during sleep, also to diagnose various types of disorders based on sleep stage analysis. There are many single-channel approaches for automatic sleep stage classification in the literature [9, 15, 22, 24, 25]. According to the available evidence [24], EEG signals are almost sufficient for reliable scoring.

To the best of our knowledge, the reported classification accuracies of the suggested methods are mostly obtained from healthy subjects [22, 25, 52]. Only a few methods in the literature are tasted on patients with various sleep disorders [23]. However, it should be noted that automatic sleep scoring methods should gain acceptable performance on analyzing EEG signals in sleep disorders. Sleep disorders (such as sleep-disordered breathing, REM behavioral disorder and sleep-related movement disorders) impose disruptive effects on the recorded signals. In these cases, sleep signals behave more irregular containing higher movement artifacts. In addition, drugs consumption may also change sleep patterns [6]. Such pitfalls may also influence both manual and automatic sleep scoring processes and the issue tends to be more profound in automatic methods. Inaccurate sleep scoring leads to misdiagnosis; consequently, the treatment based on this wrong diagnosis cause negative consequences on patients' disorder outcome and well being [6]. A few reports confirm that due to irregularity of sleep EEG among patients, the scoring accuracy do not exceed 75% which is considered below expected standards [1].

This paper reviews most of state-of-the-art automatic sleep scoring methods with their pros and cons being discussed. Such insights would be expected to help implementing several single-channel methods and apply them to normal and patient groups in order to assess the performance of published methods in different circumstances. To our knowledge, thus far, no comprehensive review on sleep EEG scoring is performed to compare the results of state-of-the-art single-channel methods on both patients and healthy subjects. Moreover, the performance of different classifiers are compared to find the best classifier for this application. In addition, to assess the robustness of these methods, Gaussian noise is added by different signal-to-noise-ratio (SNR) values and their performance are measured in presence of the noise.

Later in this report Section II explains the qualitative and quantitative sleep stage assessments. Section III describes several single-channel based methods in details. In Section IV, results of these methods on both normal and patient data are demonstrated. The final Section is dedicated to the discussion and conclusion.

2. Methodology

Sleep stages can be qualitatively/ quantitatively analyzed. In this Section, first visual sleep stage scoring criteria (qualitative methods) are explained. Then, several quantitative sleep scoring methods are introduced in detail.

2.1. Polysomnographic Data and Qualitative Assessment

Sleep medicine uses polysomnography (PSG) as an efficient method to record several biological signals to evaluate the sleep quality. PSG recordings generally involve overnight monitoring of patients sleep EEG, airflow through the nose and mouth, respiratory rate, blood pressure changes, electrocardiogram (ECG) signals, blood oxygen level, EOG signals, as well as the chin and legs surface EMGs [23, 26].

The qualitative analysis (visual inspection) of a whole night PSG recordings is performed through one of the two available standards [27] including the traditional Rechtschaffen and Kales (R&K) [28] and the more recently-developed standards laid down by the American academy of sleep medicine (AASM) [29]. Based on both the R&K and AASM criteria, EEG signal is the most informative signal compared to others. Experts analyze the EEG signals visually within successive 30 second intervals (epochs) mainly based on its standard rhythms (frequency bands). They assign a sleep stage as a label to each epoch successively [27]. The standard sleep EEG rhythms are categorized as Delta, Theta, Alpha, Beta bands (Table 1). Moreover, two important events happening through sleep EEGs are Sleep Spindles and K-complexes, where both exclusively occur in the second sleep stage.

Table 1: The Frequency Range of Sleep EEG Bands and Events

Freq. Band	Freq. Range (Hz)
Delta	0.5 – 4
Theta	4 – 8
Alpha	8 – 13
Beta	13 – 30
Sleep Spindles	12 – 14
K-Complex	0.5 – 1.5

For almost three decades, the R&K sleep classification manual was the only widely-accepted standard to describe the human sleep process [31]. According to the R&K criterion, sleep study comprises seven stages including: wakefulness (W), non-rapid eye movement (NREM) including stage 1, stage 2, stage

3, and stage 4, rapid eye movement (REM) and movement time (MT). Although the recommended setup is brief and instruction is easy to follow, many issues in sleep study are still remained unresolved [32].

Regarding the most recent AASM manual, at least 3 electrodes should be placed on frontal, central, and occipital head regions to record EEG signals [29]. The criteria concerning sleep-wake transition, sleep stages, sleep spindles, K-complexes, micro arousals, slow wave and REM sleep are revised. Unlike the R&K recommendations, in the new manual, stages 3 and 4 are merged into N3 and the MT stage is retracted from analyses. The trend transition from R&K rules to new AASM standards, left only a minor influence on total sleep time (TST), sleep efficiency (SE) and REM stage measures while it affected the quantification of sleep latency, wakes after sleep onset (WASO) and the distribution of NREM sleep stages. These result in implications both in the clinical and research fields.

According to the AASM manual, each of the the five stages is defined below and also illustrated in Fig. 1:

- W: Awake state (stage W) is characterized by alpha or faster frequency bands occupying more than 50% of the epoch, frequent eye movements and high EMG tone.
- N1: Stage N1 is scored when alpha occupies more than 50 % of epoch while theta activity, slow rolling eye movements and vertex waves are evident.
- N2: Stage N2 is scored when sleep spindles or K-complexes (less than 3 minutes apart) are noted.
- N3: Stage N3 is characterized by delta activity detected in over 20% of the epoch length.
- REM: Upon sleep scoring an epoch is marked as REM when saw-tooth waves along with rapid eye movements as well as lowest EMG signals are observed through each epoch.

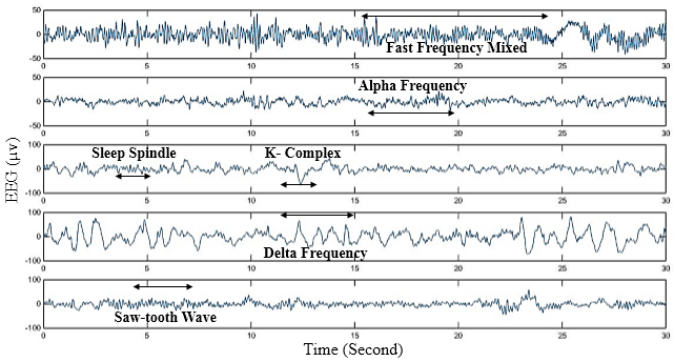

Figure 1: This figure illustrates Wake, N1, N2, N3, REM stages, from top to bottom panels, respectively.

2.2. Quantitative Assessment

Several automatic sleep stage classification methods are presented in the literature [88, 89]. To demonstrate the growing interest on sleep stage classification, the number of papers published in 28 relevant journals from the year 2000 to 2015 are shown in Table 2. Among such publications, "Computer Methods and Program in Biomedicine" and "Neuroscience Methods" are of the great interest.

As demonstrated in Fig. 2, the rate of publications has remained steady between 2000 and 2009, while it has quickly increased from 2010 to 2015. This may indicate the recent growing attention and joint cooperation of biomedical engineers and practitioners to achieve an accurate automated (expert-independent) method for sleep stage scoring.

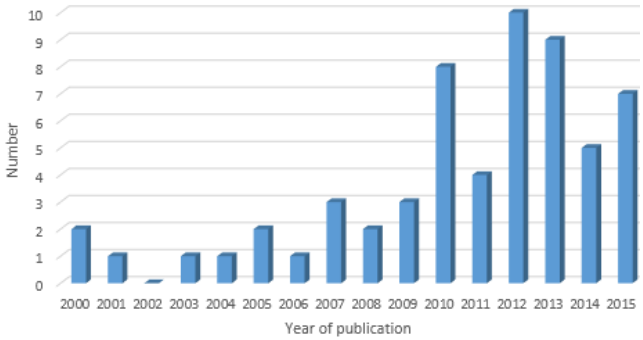

Figure 2: Distribution of research papers by year of publication from 2000 to 2015.

Based on the available literature, the quantitative sleep stage scoring schemes comprises 3 common steps include preprocessing, feature extraction and classification. Furthermore, in some papers, feature selection step is added following feature extraction in order to find suitable subset of features [54]. What follows is introducing more frequent techniques in each step. The process for automatic sleep stage scoring is schematically illustrated in Fig. 3.

2.2.1. Preprocessing

The presence of artifacts might lead to the misinterpretation, inaccuracy and distorted quantitative results; therefore, a preprocessing step is necessary to remove artifacts and magnify informative components of raw EEG signals prior to any further analysis [65]. There are biasing factors which potentially affect the accuracy of sleep EEG scoring [23]. These factors can be measurement noise, physiological factors (including psychophysiological variants and aging) and pathological conditions (such as disordered sleep breathing and REM behavioral disorder) [29, 30].

Despite hardware filtering during the recording, EEG signals are still contaminated by other noise sources [37]. Diffuse artifacts may mask cerebral activity and simulate sleep phasic events (sharp vertex waves, K-complexes). Moreover, some noises and artifacts are created by the measurement apparatus. The power line interference (the 50/60 Hz components) or movements of the wires and electrodes can impose

Table 2: Distribution of Research Papers by Journal.

Journal Title	**Amount**	**Percentage** (%)
Computer Methods and Program in Biomedicine	6	10.27
Journal of Neuroscience Methods	6	10.27
Computers in Biology and Medicine	5	8.47
Journal of Medical Systems	5	8.47
Expert System with Application	4	6.78
Artificial Intelligence in Medicine	3	5.08
IEEE Transactions on Biomedical Engineering	3	5.08
Biomedical Signal Processing and Control	2	3.39
International Journal of Neural System	2	3.39
Journal of Sleep Research	2	3.39
Methods of information in medicine	2	3.39
Neurocomputing	2	3.39
Sleep Medicine Reviews	2	3.39
Advances in Adaptive Data Analysis	1	1.69
Biomedical Engineering Applications, Basis and Communications	1	1.69
Biomedical Engineering On-line	1	1.69
Brain Research Bulletin	1	1.69
Clinical Neurophysiology	1	1.69
Computers and Biomedical Research	1	1.69
Frontiers of Computer Science	1	1.69
IEEE Transactions on Neural Systems and Rehabilitation Engineering	1	1.69
IEEE Transactions on Instrumentation and Measurement	1	1.69
International Journal of Adaptive Control and Signal Processing	1	1.69
Medical and Biological Engineering and Computing	1	1.69
Neural Computing and Applications	1	1.69
Sleep	1	1.69
Sleep Medicine	1	1.69
World Academy of Science, Engineering and Technology	1	1.69
Total	**59**	**100**

changes in the baseline characteristics of signals. Moreover, some artifacts are originated from non-cerebral sources such as eye movement, muscle and cardiac sources. Rolling eye movements happen upon the consciousness while in REM sleep the eye movements acquire a saccatic patterns. In addition, give the comparatively high amplitude level of ECG signals, the QRS complex regularly interfere in the EEG signals causing spiky pattern [33]. Muscle activities generate EMG signals which can corrupt and distort EEG signals, especially during the arousal periods [23, 34]. Further to the instrumental and biological artifacts, inter-individual variability of background activity and phasic events as well as age-related changes may influence the sleep stage scoring. On the other hand, focal cerebral lesions might decrease phasic event's amplitude, density and therefore mimic the REM pattern. In patients with wake or nocturnal epilepsy, the interictal EEG epileptic activity may similarly mask sleep activity or bias sleep EEG phasic events [6]. Consequently, the above biasing factors may hinder accurate scoring and emphasize the role of preprocessing in order to avoid deceiving clinical results both in visual and automatic scoring.

Several methods are proposed in order to attenuate or eliminate the effects of artifacts in EEG signals. These methods are described below:

- **Digital Filtering**: It covers a vast class of different filters which can be linear or non-linear. To pre-process raw EEGs, it is common to apply a band-pass filter (e.g. butterworth) to eliminate the undesired parts. In order to eliminate the power line noise from the signals, a digital Notch filter is utilized [37]. Band-pass filters [39] such as Butterworth [37] are used to reduce the muscle artifacts and eliminate linear trends [36]. Further to the linear filtering, adaptive filters are used to remove the effects of the EOG and ECG artifacts in which the frequency contents overlaps with EEG spectrum [41].

- **Regression**: Regression methods are able to learn a signal behavior. If an additive color noise distorts the EEGs, a regression method is used to learn this trend and therefore by modeling the color noise, it can be subtracted from the recorded signal. [42].

- **Wavelet Transform (WT)**: WT decomposes a signal into its sub-bands. By calculating the energy in each sub-band and applying a threshold, the noisy and undesired bands are eliminated [47]. Since there are many mother wavelet

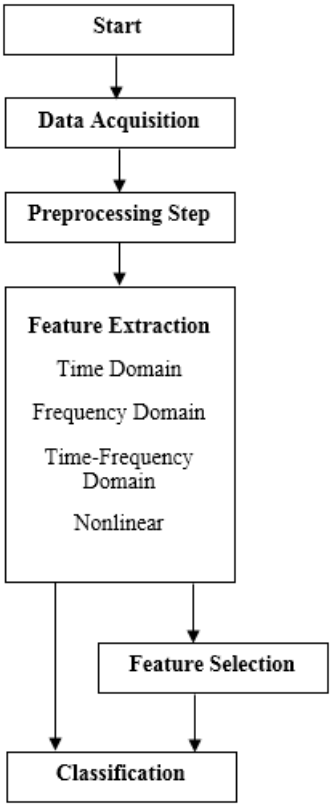

Figure 3: Automatic Sleep Stage Classification Process.

functions, the one with highest similarity to the nature of EEG (in terms of frequency bands) should be selected. Reviewing the existing papers, we can observe that Symmlet and Daubechies mother wavelets are the most common used wavelets for EEG processing [87].

- **Blind Source Separation (BSS)**: An alternative artifact removal approach is BSS algorithms. Independent Component Analysis (ICA) [33, 34]is one of the well-known BSS methods to produce de-correlated brain source signals after fading the artifacts. By calculating the energy of source signals and using a threshold, the independent signals with lower energy level will be eliminated. Finally, with an inverse transform, the denoised signals will project back into the spatio-temporal domain.

After the preprocessing step, based on the mentioned criteria, EEG records are segmented into the successive 30 sec epochs with no overlap, after which, further analysis are carried out at each epoch.

2.2.2. Feature Extraction

Features are descriptive values that project the content of a signal from a special aspect. Usually, different types of features are extracted from a windowed signal; therefore, to avoid collecting redundant information, the elicited features should be as independent as possible. Extracting informative, discriminative and independent features is a crucial step to prepare a suitable set of values for a classifier [48]. A wide variety of signal processing techniques are utilized to extract discriminative information from the EEG signals in order to automatically classify the sleep stages [10, 65]. Features can be totally divided into four categories (Time domain, Frequency domain, Time-Frequency domain, Nonlinear) which are briefly discussed below:

Time Domain Features. Time domain features can represent the morphological characteristics of a signal. They are simply interpretable and suitable for real-time applications. Some of the widespread time-base features are described below:

- *Statistical Parameters*: 1st to 4th-order moments of time series i.e., mean, standard deviation, skewness and kurtosis, respectively [39, 52], median and 25th, 75th percentile of the signal distribution [54] are known as the most effective time domain features for EEG signals. Their mathematic formulas are presented in Table 3. They are not only applied to EEG signals but also used to extract features from the other biological signals including EOG and EMG [13, 37].

- *Hjorth Parameters*: Hjorth parameters are introduced by Bo Hjorth in 1970 [38]. These parameters include Activity (H_a), Mobility (H_m) and Complexity (H_c) as the measure of variance of a time series (x), proportion of standard deviation of the power spectrum and change in the frequency, respectively. They are widely applied features for EEG analysis [39, 40] and described below.

$$H_a = var(x(t)) \tag{1}$$

$$H_m = \sqrt{\frac{var(x(t)\frac{dx}{dt})}{var(x(t))}} \tag{2}$$

$$H_c = \frac{H_m(x(t)\frac{dx}{dt})}{H_m(x(t))} \tag{3}$$

- *Zero Crossing Rate (ZCR)*: ZCR represents number of times that a signal crosses the baseline (Eq. 4). The baseline can be obtained by taking a mean value from a windowed signal [39]. It measures the number of sign-changes through a windowed signal. Given the differences in quality and the appearance of the each sleep stage in the time domain, ZCR may vary from one stage to another; as such, it can be used as a suitable feature for sleep stage classification [52, 54, 55]. Nevertheless ZCR is highly sensitive to additive noises.

$$ZCR = count\{n|(x_n x_{n-1}) < 0\} \tag{4}$$

Frequency Domain Features. Frequency domain features are versatile features which are repeatedly utilized for describing changes in EEG signals [7, 14, 35, 37, 58]. There are several approaches to extract frequency features, herewith summarized.

Table 3: Time Domain Features.

Feature Name	Formula	Elucidation
Mean	$\mu = \frac{1}{N}\Sigma_{n=1}^{N} x_n$	x_n $(n = 1, 2, ..., N)$ is a time-series.
Variance	$var = \frac{1}{N-1}\Sigma_{n=1}^{N}(x_n - \mu)^2$	A measure of how far a set of numbers is spread out from mean.
Standard Deviation	$std = (Var)^{1/2}$	A measure of dispersion of a dataset.
Skewness	$skew = \frac{E[(x-\mu)^3]}{(E[(x-\mu)^2])^{3/2}}$	A measure of symmetry of a dataset about its mean.
Kurtosis	$kurt = \frac{E(x-\mu)^4}{(E(x-\mu)^2)^4}$	A measure of tailedness of a dataset.
yth Percentile	$ythper = \frac{yN}{100}$	A sample number below which a given percentage of a dataset fall.
Median	M= $\begin{cases} \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2} & \text{for even N} \\ x_{\frac{N+1}{2}} & \text{for odd N} \end{cases}$	A sample number separating higher from lower half of a dataset.

Spectral Estimation To obtain spectral characteristics of EEG signals, first, time series should be transfered to the frequency domain. Due to the stochastic nature of EEG signals, signal autocorrelation needs to be estimated first. Then, by taking Fourier transform (FT) from the autocorrelation function, Power Spectral Density (PSD) is achieved. The two approaches to estimate PSD of a windowed signal include parametric and non-parametric methods.

- *Parametric methods*: parametric methods are model-based approaches in which the spectrum is estimated by the signal model. Autoregressive (AR), Moving average (MA) and Autoregressive moving average (ARMA) are amongst popular methods in EEG processing [49]. When signals have low SNR and long length, parametric approaches present a smoother and better estimation of power spectra. However, model order determination of AR is a main concern in such methods, since AR is able to model the two others. To solve this problem, some criteria such as Whilst Akaike's information criterion (AIC) [50] are introduced in the literature [51].

- *Non-parametric methods*: In non-parametric approaches, PSD values are calculated directly from the signal samples in a given windowed signal. They are based on Fourier transform [81]. Among non-parametric methods, Periodogram and Welch [82] schemes are general for calculating PSD and extracting features from EEGs. Easy implementation based on Fast Fourier Transform (FFT) algorithm is the merit of non-parametric methods; however, they gain low frequency resolution when the signal interval becomes short.

Higher Order Spectra (HOS) : HOS is another method used to extract frequency domain features. HOS techniques are employed in several biomedical applications [72]. HOS represents the frequency content of higher order statics of signals (cumulants) [75, 78]. Power spectrum equals to second-order spectra which loose the phase information. Also, the 3rd order spectrum is known as bi-spectrum which is showed in Eq. 5.

$$B(f_1, f_2) = E[X(f_1)X(f_2)X^*(f_1 + f_2)] \quad (5)$$

where $E[.]$ is expectation operator, $X(f)$ is the Fourier transform of the signal and f is the frequency.

The advantage of using HOS is its ability to reveal non-Gaussian and nonlinearity characteristics of the data [78, 72]. Since EEG is a complex signal subject to some non-linear interaction in its frequency components, HOS methods are useful approaches in analyzing sleep EEG signals [75].

Time-Frequency Domain Features. Due to the non-stationary nature of EEG signal, i.e. variation in its properties during the time, a wide rang of time-frequency methods are utilized as an efficient tool in this field. There are three ways to transfer a signal into the time-frequency plane including signal decomposition, energy distribution and modeling. The first two ways are utilized in the sleep applications [80, 88] (explained below).

- *Signal Decomposition*: These methods decompose signals to a series of basis functions. Short Time Fourier transform (STFT) is a very simple time-frequency analysis. In this method, first, signal is uniformly windowed and then FT is applied to each window as described in Eq.6.

$$STFT_x(t, f) = \int_{-\infty}^{+\infty} x(\tau)h^*(\tau - t)\exp(-j2\pi f\tau)\, d\tau \quad (6)$$

where $h(.)$ is a window function.

To use this method as a mean of feature extraction, the issue of time and frequency resolution need to be well noted. There is an inverse relationship between the length of the window and the time resolution. In addition, the time and frequency resolutions have an trade-off together. For instance, the shorter the data segment, the higher the time

resolution will be achieved and therefore frequency resolution decreased. However, a fixed specific length of the window is considered to determine STFT leading to a limited frequency resolution.

Wavelet Transform (WT) is a popular time-frequency transformation that applies different filters to decompose a signal into dyadic frequency scales. These filters are generated from a function which called mother wavelet. Both discrete and continues forms of WT are utilized for sleep classification [79, 88]. Continuous Wavelet Transform formula of a given signal of $x(t)$ is shown in Eq. 7.

$$W_{a,s} = \int x(t)\phi_{a,s}(t)dt \tag{7}$$

where a and s are scale and time shifting parameters, respectively. $\phi_{a,s}$ is mother wavelet transformation which is defined as:

$$\phi_{a,s}(t) = \frac{1}{\sqrt{|a|}}\phi(\frac{t-s}{a}) \tag{8}$$

WT is a powerful method since it describes the signal into different frequency resolutions where the decomposed signals are orthogonal in most mother wavelets. In addition, a colored noise cannot invade all of the wavelet features in different scales [87].

- *Energy Distribution (ED)*: ED represents the distribution of energy simultaneously through both time and frequency domains. Several methods are placed in this category. Choi-Williams distribution (CWD), is energy conservative which maintains time shifts as well as the frequency shifts. These characteristics make it a favorable method for the time-frequency analysis [88, 90].

  Wigner-Ville distribution (WVD) is another non-linear time-frequency method which is widely used to analyze non-stationary signals. However, the presence of cross-terms in the time-frequency domain for multi-component non-stationary signals hinder its efficiency. Therefore, the smoothed pseudo WignerVille distribution (SPWVD) (the modified version of WVD) is proposed to avoid the effects of cross-terms [80], as described in Eq. 9.

$$S(t,f) = \int h(\tau)[\int f(u-t)x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})du]exp(-jw\tau)d\tau \tag{9}$$

  where f(.) and h(.) are the time and frequency smoothing functions, respectively.

  In addition, Hilbert-Huang Transform (HHT) is a new method in analyzing nonlinear and non-stationary signals such as EEGs. It is recently employed in many applications in biomedical fields including sleep stage scoring [88].

Non-linear features. Since EEG signals exhibit non-linear characteristics and complex dynamics [43], non-linear measures are used as effective techniques for EEG processing [44, 61]. They are also widespread in sleep stage scoring to extract efficient features from sleep EEG [15, 18, 44, 56, 57]. Non-linear methods are mainly divided into two categories described as follow:

- *Entropy and Complexity-based*: Entropy-based methods calculate the irregularity and impurity of a signal in the time domain [61]. In other words, the more the changing patterns inside a windowed signal get regular, the less would be the entropy value of that windowed signal and vise versa. The most famous entropy measure is introduced by Shannon as described below:

$$Entropy(X) = -\Sigma_{i=1}^{N} P_i Ln(P_i) \tag{10}$$

  where N is the number of samples and P_i is the probability of i^{th} sample.

  Several entropy and complexity measures such as *Renyi's Entropy* [88], *Sample Entropy* [7], *Tsallis Entropy* [43], *Permutation Entropy* [89], *Lempel-Ziv* [39], Multi Scale Entropy (MSE) [20] and *Approximate Entropy (ApEn)* [56] which are introduced in the literature. They have enough potential to serve as informative features in sleep stage classification [10].

- *Fractal-based*: Constructing a noisy-like shape can be done by repeatedly applying an operator on a geometric element. Several phenomena exist which have a noisy behavior while they are rule-base in nature. For instance, the behavior of EEG signals or shape of leaves in a tree look random while one can hardly claim that creating EEG signals or leaves in a tree are random. For the first time a novel method is proposed by Mandelbrot [73] to measure the fractal dimension in of irregular shapes. The concept of fractal dimension can describe the behavior of random-like shape by determining the amount of self-similarity on that given shape or signal. There are many fractal-based methods like *Correlation Dimension* [75], *Lyapanov exponent* [39] and *Hurst exponent* [37] which first map a signal into the phase space and then measures the self-similarity of its trajectory shape. In other words, instead of analyzing a signal in the time domain, these methods analyze its trajectory behavior in the phase space, in terms of self-similarity. There are some other methods determining the roughness or irregularity of the signal in the time domain such as *Katz* [53], *Petrosian* [54] and *Higuchi fractal dimension (HFD)* [39]. Experimental results from different studies have suggested that measuring the behavior in the phase space is more accurate than the same in the time domain, though it imposes a higher computational burden [53].

2.2.3. Feature Selection

In some sleep stage classification methods especially in multi-channel approaches, feature selection techniques are applied after the feature extraction stage to find discriminative subset of features. The purpose of adding feature selection part is to obtain minimum number of features without redundancy

while producing higher accuracy. Furthermore, it avoids the issue of over-fitting and reduce the computational time. Since an overnight PSG record is large, feature selection is useful especially once different types of features are extracted.

To this end, some statistical techniques are proposed in the literature. These techniques try to find appropriate and discriminative subset of features to distinguish different sleep stages. Sequential forward selection (SFS) [15] and sequential backward selection (SBS) [15, 45] are two simple strategies to explore through the features and find a proper subset. Nevertheless, SFS and SBS highly suffers from the lack of backtracking. In addition, feature extraction algorithms such as Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) [46] are used for feature reduction. On the other hand, meta-heuristic algorithms like Genetic Algorithm (GA) [61] and Particle Swarm Optimization[62] are repeatedly used for feature selection.

2.2.4. Classification

Various classifiers are utilized to classify the elicited features and assign a sleep stage into each epoch. These classifiers are learned to construct linear/non-linear boundaries to separate feature vectors of different classes. Some of the popular classifiers used in sleep stage scoring are herewith explained.

- *K-Nearest Neighbor (KNN)*: This is a local classifier which is suitable for multi-modal distribution data. In other words, when the samples of a certain class are modularly scattered in different locations of the feature space, even flexible and strong classifiers cannot accurately put a boundary around this localities. KNN assigns a label to an input patterns based on the majority vote of its k-nearest samples [82].

  It can be statistically demonstrated that when k=1, the error probability of 1NN is bounded between that of the optimum classifier (Bayes) and twice of this value [83].

$$P_{error}(Bayes) \leq P_{error}(1NN) \leq 2P_{error}(Bayes) \qquad (11)$$

- *Support Vector Machine (SVM)*: For the first time, a generalized classifier trying to minimize the risk of error instead of minimizing the classification error [60] is proposed by Vapnic [59]. SVM considers a margin around its hyperplane while other conventional classifiers just attempt to form a boundary, weather linear or nonlinear, between two classes. In fact, SVM tries to maximizing the margin width simultaneous to minimizing the classification error of samples, within that margin. Therefore, finding the separating plane of SVM needs to solve the constrain optimization problem. Since the objective function of SVM as describe in Eq. 12 is convex, SVM will be a stable classifier in terms of boundary learning. Since SVM optimization formula has a constraint, Lagrange coefficient is inserted into its objective function; whereby in each sample a Lagrange coefficient is determined.

$$L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \phi(x_i) \phi(x_j) \qquad (12)$$

where α_i is the Lagrange multiplier of the i^{th} sample, $\phi(.)$ is the kernel function, x_i is the i^{th} input and y_i is its corresponding label.

The values of Lagrange coefficients belong to the samples located within the margin space (support vectors) are positive values bounded within:

$$0 \leq \alpha_i \leq c \qquad (13)$$

Where c is a user-defined parameter.

In contrast, other samples (majority of instances) located outside of the margin space do not have any role in determining the SVM hyperplane, since their Lagrange coefficients become zero. The boundary of SVM is determined by

$$W = \Sigma_{i=1}^{N} \alpha_i y_i x_i \qquad (14)$$

where N is the number of samples.

- *Random Forest (RF)*: RF is proposed by Breiman [64], consists of an ensemble of tree-structures. Each individual tree depends on a random vector sampled values and plays the role of an individual classifier. The main difference of RF to the other classifiers is that feeding the input samples to the trees are performed as random as possible through a random selection followed by different bootstrap selections. This process are repeated several times in order to blend the samples in order to desensitize the effect of noisy and outliers samples of the training phase. Finally, the output is determined by voting of the trees' outputs. The general structure of RF classifier is summarized in Fig. 4

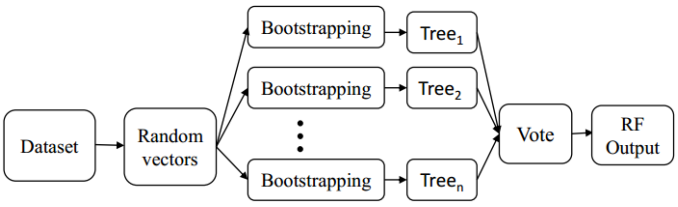

Figure 4: RF Structure

- *Linear Discriminant Analysis (LDA) & Nearest Centers (NC)*: LDA is developed by Fisher in 1936 and optimized by a very popular criterion function, called as Fisher criterion [63]. In two-class problems LDA can be considered as a classifier whereas in multi-class problems, LDA is acted as a feature-extraction method. Since sleep EEG contains 5 classes, LDA provides more separable features

for the next classifier. LDA tries to maximize the ratio of between-to-within classes' scatter matrices. This is done by projecting input samples onto a few number of hyperplanes (depends on the number of classes) such that the separability among the samples is maximized in the projected space [61]. The Fisher criterion is described as:

$$J(W) = \frac{W^T S_B W}{W^T S_W W} \tag{15}$$

where W is potentially a hyperplane (in two-class problems) or a matrix of hyperplane (in multi-class problems).

Incidentally, S_W and S_B are between and within class scatter matrices, defined as:

$$S_{W_i} = \Sigma_{x \in c_i}(x - m_i)(x - m_i)^T, \quad i = 1, 2, ..., c_i \tag{16}$$

$$S_W = \Sigma_{k=1}^{c} S_{Wi} \tag{17}$$

where x is the input vector, c is the number of classes and m is mean value.

$$S_B = \Sigma_{k=1}^{c}(m_k - m)(m_k - m)^T \tag{18}$$

The final decision is made by applying a distance-base classifier (NC) to the LDA outputs. In other words, LDA acts as a feature extractor and the projected features are assigned to the corresponding classes according to the minimum distance of the project sample to the center of each class, separately [87].

- *Neural Network (ANN)*: ANN is an artificial information processing system comprising a number of interconnected processors called neurones. Scientists tried to assign functions for these neurons following the real behavior of biological neural cells. The most employed functions of neuron in different ANNs are sigmoid, hyperbolic and linear functions. ANN architecture can contain several hidden layers each includes several neurons. In most recent employed ANNs, just one hidden layer is used to avoid the over-fitting phenomenon. According to the statistics, most of the utilized neural networks include an input layer, a hidden layer and an output [65]. The number of nodes in input layer depends on the number of input features and the number of nodes in the output layers depends on the number of classes. The number of neurons in the hidden layer can be determined through the cross validation phase. ANNs are widely used in sleep stage scoring [7, 16, 13, 65, 66]. The schematic diagram of a feed-forward multi-layer perceptron (MLP) is shown in Fig 5.

  Outputs (y) are calculated based on a transfer function as illustrated in Eq. 19.

$$y_j = \Sigma_{k=1}^{t} W_{kj} f_k \tag{19}$$

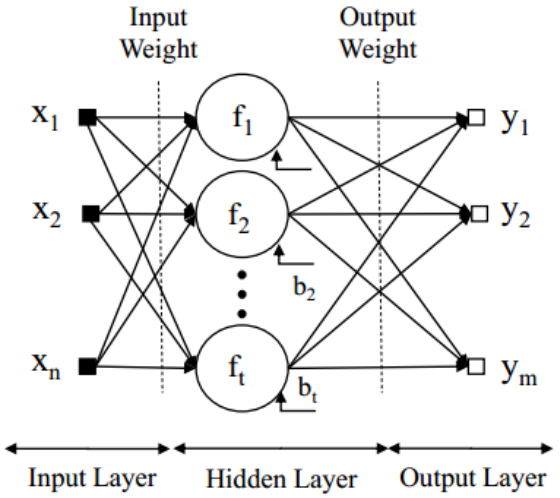

Figure 5: ANN architecture, feed-forward multi-layer perceptron (MLP).

$$f_k = f(W_{ik} x_i + b_k) \tag{20}$$

where W is the weight matrix and b is a certain value to control the on/off threshold of neurons which are adjusted during the training procedure.

In contrast to the sluggish training phase of ANNs, these networks act very fast in the test phase. The main drawback of ANNs is their high tendency of being over-fitted to the noisy samples which are mostly located along the margin space between the classes. ANNs do not include a margin to consider a little space for deterioration of test samples around its boarder. However, the selection of its parameters, such as network size and training algorithms, is trivial since such parameters can affect the classifier performance.

- *Gaussian Mixture Model (GMM)*: This statistical model is utilized in a wide range of applications as the density estimators [74] and classifiers [75]. This model consists of several Gaussian functions integrated with different coefficients and finally the weighted summation on these probabilities construct the output [75]. Therefore, it can be said that a GMM model has 3 types of parameters which need to be estimated separately, i.e. mean vector (μ), covariance matrix (Σ) and weight (W) for each Gaussian, separately [76]. Schematic diagram of GMM along with its corresponding formulas to determine those parameters for a sequence of T training vectors $X = \{x_1, ..., x_T\}$ and m component of densities are outlined below.

$$\lambda = \{\mu_i, \Sigma_i, W_i\}, \quad i = 1, 2, ..., m \tag{21}$$

$$W_i = \frac{1}{T}\Sigma_{t=1}^{T} p(i|x_t, \lambda) \tag{22}$$

$$\mu_i = \frac{\Sigma_{t=1}^{T} p(i|x_t, \lambda) x_t}{\Sigma_{t=1}^{T} p(i|x_t, \lambda)} \tag{23}$$

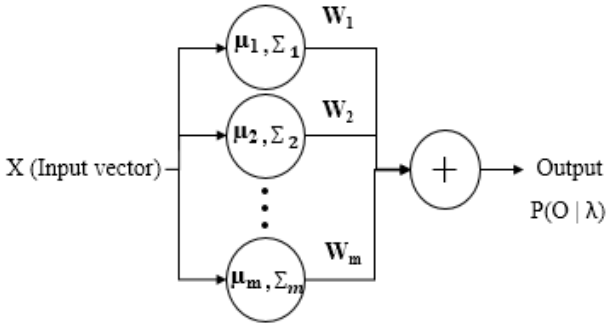

Figure 6: GMM Structure

$$\sigma_i^2 = \frac{\Sigma_{t=1}^{T} p(i|x_t, \lambda) x_t^2}{\Sigma_{t=1}^{T} p(i|x_t, \lambda)} - \mu_i^2 \tag{24}$$

where σ_i^2 is variance. It should be noted that when the two variables are identical the covariance is equal to variance.

In fact, the independent Gaussian functions with different weights can model every arbitrary distribution functions. In order to use GMM as a classifier, the distribution of input features of each class is learned by a GMM where each GMM acquires a specific label. Consequently, when a sample is entered to the parallel GMMs, its label is assigned according to the label of the given GMM which produces the highest probability.

- *Hidden Markov Model (HMM)*: HMM is a comprehensive version of Markov chain model in which the sequence of states is unknown. HMM is dynamic classifier which can tolerate the time warping of input observations, $O = \{O_1, ..., O_T\}$ where O_i is the i^{th} observation (feature vector). The semi-continuous HMM consists of a few continuous states, each constructed by Gaussian functions, which are connected by discrete transition probabilities. The main structure of HMM is the semi-continuous left-to-right model, i.e.

$$\lambda = (\pi, b_i, A) \tag{25}$$

Where A is the transition matrix, b_i is the i^{th} state probability and π demonstrates the probability of the first state.

HMM training process involves 3 stages. First the HMM parameters should be determined by the Baum-Welch algorithm. Afterwards, the sequence of states is found by Viterbi algorithm [14]. Following the above, the problem is simplified into calculating the output of Markov model (stage 3) [14]. The number of states and their number of Gaussian functions are determined by the Expectation Maximization (EM) criterion [77]. Fig. 7 shows a schematic diagram of HMM.

Where a_{ij} shows the probability of moving from i^{th} to jth state.

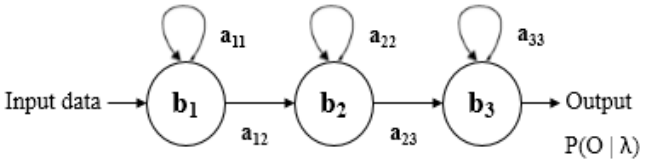

Figure 7: HMM architecture.

- *Clustering*: Clustering is a kind of grouping or unsupervised categorization where the samples lack labels. Therefore, this grouping should be performed based on a similarity measure such as distance metrics, information criteria and statistical measures. The idea of clustering is to create different clusters such that each cluster contain similar samples according to one of the mentioned criteria. Clustering methods are generally divided into four categories including flat (e.g. K-means) [67, 68], hierarchical (e.g. single and complete-linkage) [69, 70], graph-base (e.g. Shared Nearest Neighbor (SNN)) [71] and density-base (e.g. DBSCAN, OPTICS and DenClue) [69].

  Clustering methods, specially K-means, are repeatedly used in sleep stage scoring [82]. In other words, after clustering the whole sleep features, it is expected that feature vectors belonging to each sleep stage are gathered in a certain cluster.

3. State-of-the-art Studies in Automatic Sleep Stage Classification

Each of the available automatic sleep classification methods is evaluated based on a certain dataset with its own signal quality, while changing in signals quality leads to deviate in the final results. In practice, due to the importance of accurate sleep scoring for patients with sleep abnormalities, these methods should have high performance in patients' recorded signals. To the best of our knowledge, evidence on various methods on a specific dataset acquired from patients with sleep disorders are thin. The present review is an attempt to discuss the most recent automatic sleep stage classification methods using single-channel EEG data.

3.1. Polysomnographic datasets

Two EEG datasets which are available on-line in the data repository physionet [84] have been employed in this study. The first dataset contains EEG signals belong to healthy subjects whereas the second one is acquired from patients who suffer from REM behavioral disorder (*RBD*). The employed datasets in this study are described as follow.

3.1.1. Sleep–EDF Dataset [extended]

This dataset includes the collection of 61 PSG data with respective sleep stages annotations [85]. It consists of 2 files including SC and ST, where the first contains PSGs of 20 healthy males and females aging between 25-34 years old, without any medication related to sleep. The records consist of EEG (from Fpz-Cz and Pz-Oz electrodes location), EOG, submental chin

EMG and event marker. The EEG signals are acquired with the sampling rate of 100 Hz. The ST file consists of 22 healthy male and female records with only mild difficulty falling asleep. In addition hardware filtering is applied to the signals. Each of the 30 sec epochs is manually scored based on the R&K standard as W, S1, S2, S3, S4, REM, MT and Unscored with digit codes of 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. 10 records from SC files are selected and the bipolar records from Pz-Oz channels as a single-channel EEG.

3.1.2. CAP Sleep Dataset

It contains PSG recordings of 108 subjects including healthy subjects as well as patients with seven various sleep disorders datasets [86]. They consist of at least 3 EEG channels (F3 or F4, C3 or C4 and O1 or O2, refer to A1 or A2), ECG, submental muscle EMG, bilateral anterior tibial EMG and respiration signals. EEG signals are sampled at 512 Hz. Prefiltering also applied to the signals including: LP (30 Hz), HP (0.3 Hz) and Notch filter (50 Hz). In addition, sleep stages annotations are available as: W=wake, S1-S4=sleep stages, R=REM, MT=body movements. In this study, 10 single-channel EEG records of *RBD* dataset are selected as the patient dataset.

3.2. Comparative Studies

1. Informative features from the higher order spectral (HOS) space are elicited for sleep stage classification by Acharya *et al.* [75]. First, bi-spectrum of EEGs were calculated in each epoch as explained in Eq. 5. Then, following features were extracted in the biospactrum space.

- *Normalized Bispectral Entropy (BE 1)*:

$$BE1 = -\Sigma_{\Omega} p_n log p_n \tag{26}$$

where

$$p_n = \frac{|B(f_1, f_2)|}{\Sigma |B(f_1, f_2)|} \tag{27}$$

Ω is the non-redundant region in Fig. 8

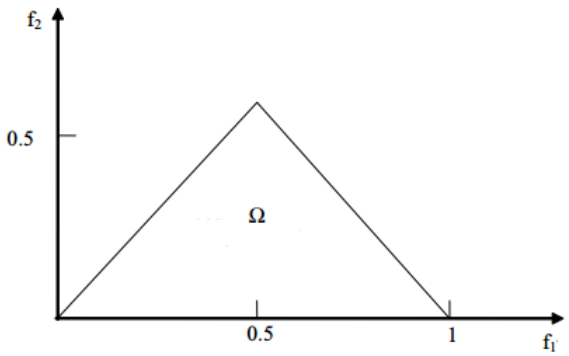

Figure 8: Non-redundant region for biospectrum calculation.

- Sum of logarithmic amplitude of bi-spectrum:

$$H1 = \Sigma_{\Omega} log(|B(f_1, f_2)|) \tag{28}$$

- First-order moment of amplitude of diagonal element in bi-spectrum:

$$mAmp = \Sigma_{\Omega} klog(|B(f_1, f_2)|) \tag{29}$$

- Weighted Center of bi-spectrum (WCOB):

$$WCOB_x = \frac{\Sigma iB(i, j)}{\Sigma B(i, j)} \tag{30}$$

After extracting the above HOS features from each epoch, a GMM classifier is trained to learn the feature vectors of each sleep stage. At last, for the final classification of a whole sleep, the elicited feature vector of each epoch is entered to all trained GMMs. Since each GMM has a label, the label of the GMM which produces the highest probability for each feature vector is assigned to the input.

2. In another study, CWT is used to represent EEG signals into the time-frequency domain [87]. To have a rich set of features, they employed three different mother wavelets including the reverse bi-orthogonal (rbio3.3), Daubechies (db20), and Gauss of order 1 (Gauss1) with center frequencies of 0.43, 0.67, and 0.20, respectively. After passing the EEGs through different wavelets' filters, entropy of each filtered signal is determined according to Eq. 31 for each of the frequency bands shown in Table 1. Moreover, Beta band is divided into two sub-bands with frequencies of $13 - 22$ and $22 - 35Hz$; therefore, 7 frequency bands are totally considered.

$$Ent = -\Sigma_{i=1}^{n} p_i log p_i \tag{31}$$

where p is the histogram distribution of wavelet coefficients in each band with n bins.

These entropy values are arranged in a feature vector for each epoch. Therefore, by calculating the entropy in seven frequency bands for each CWT separately, a feature vector consists of 21 elements for each epoch is formed and continued for the entire EEG signals. The sleep stage classification process is ended by feeding LDA+NC classifier with the extracted features.

3. The Welch method is employed to extract 129 features from each epoch by *Günes et al.* [82]. Due to the high number of input features, four new features including minimum, maximum, mean value and standard deviation are extracted from each feature vector. Then, the feature weighting process is pursued using the K-means clustering based feature weighting (KMCFW) as described in Algorithm 1.

Algorithm 1 K-means clustering based feature weighting (KMCFW)

Require: the dataset including classes

1: Find centers of features using K-means Clustering (KMC)
2: Calculate the ratios of means of features to their centers
3: Multiply these ratios with each feature

Afterwards, these weighted features are applied to KNN classifier, where K is set to 30 through the cross validation phase.

4. In another attempt, MSE feature containing 1 to 13 scales along with AR coefficients (order 8) are extracted to describe the EEG behavior by Liang *et al.* [20]. MSE algorithm is explained below. It should be noted that AR coefficients are extracted from the delta band of EEG signals.

Algorithm 2 MSE algorithm

Require: A time series $x = \{x_1, x_2, ..., x_N\}$

1: divide x into non-overlapping windows of length τ (scale).
2: Determine the average of each window as an one dimensional sample.
3: Identify the Entropy of these successive samples as a sample entropy feature.

Therefore, the combination of 8 AR coefficients and 13 MSE features build the feature vector with 21 elements in each epoch. The process is developed by classifying the extracted features using the LDA+NC classifier.

5. In the last investigated method, three different time-frequency techniques including CWT, CWD and HHT, are separately adopted to represent EEG content in the time-frequency domain [88]. Each of the time-frequency transforms divides the EEG into dyadic sub-band (as described in Table 1). To elicit informative features, *Renyi's entropy* is applied to each sub-band within each epoch. The formula of Renyi's entropy is described in Eq. 32.

$$REnt = \frac{1}{1-\alpha} log_2(\Sigma_{i=1}^{n} P_i^{\alpha}) \tag{32}$$

where P_i is the histogram distribution belonged to the coefficients of each mentioned time-frequency algorithms in each band with n bins. α is the order of Renyi's entropy. When α equals to 1, the *Renyi's Entropy* converge to *Shannon Entropy*.

These entropy values are extracted from the sub-bands of each time-frequency transforms, separately. In their proposed method an assessment is carried out by applying the entropy values, belong to each transform, to a random forest (contains 10 trees) classifier. According to the obtained results, the CWT feature provides the highest accuracy, compared to the others.

4. Results

After investigating the results from five elegant studies, their applied methods are applied to the present datasets (containing healthy subjects and patients) as described in Section 3. The employed datasets are annotated by experts according to R&K rules [85, 86]. Since AASM is an improved and more recent standard for sleep stage scoring, to match the results with this standard, S3 and S4 are merged into one stage (N3). These five methods are evaluated in terms of classification accuracy and their robustness against additive noise (with different SNRs). To find the best combination of features and classifier for achieving accurate and reliable results, here, additional implementations are suggested and executed by applying the deployed features by the mentioned papers to different classifiers including: RF, KNN, LDA+NC, GMM and multi-class SVM.

To proceed to the preprocessing phase, the raw signals are fairly cleaned by applying a bandpass Butterworth filter (order 5 with frequency cutoff between $0.5 - 35Hz$) to remove the base-line drift and linear trends. Next, EEGs are segmented into successive 30 second epochs and then the features and classifier of each paper are separately executed to those epochs in order to produce the results on both groups. The accuracies of each sleep stage classification method are demonstrated in Table 4. Results have emerged from a cross validation phase in which features vector of one subject is considered as the test set and the rest of subjects are considered as the train sets. This substitution for constructing new train and test sets is continued in parallel up to the number of subjects. This cross validation is called leave-one(subject)-out in which the train and test sets of subjects are selected totally blinded to each other.

Since the main concern of automatic methods is the accurate detection of the sleep stages for abnormalities in patients records, to make a real evaluation among these papers on real patients, here the methodologies of the papers are applied to both normal and patient groups. Not surprisingly, lower performance on patients is expected compared to the normal groups. Fig. 9 illustrates the classification accuracy of sleep stages on both groups. The results in Fig. 9 are presented as an average for all sleep stages, while Table 4, illustrates the classification accuracy of those methods on just normal subjects upon the entire sleep, and each sleep stage separately. Moreover, a short description of each method is presented in Table 4.

The best performance in terms of accuracy belongs to the 5^{th} approach in which entropy of CWT and RF are chosen as the features and classifier, respectively. Apparently, this approach provide over 80% classification accuracy in average over the whole sleep stages. Focusing on the classification accuracy of each stage, one can note that the best and worst results belong to the awake and N1 stages, respectively.

Moreover, as demonstrated in Fig. 9, the accuracy of five methods on the *RBD* dataset is reduced more than 25% in average. Although the accuracy of all methods has decreased for the patients' dataset, the 5^{th} method again presented the best classification performance.

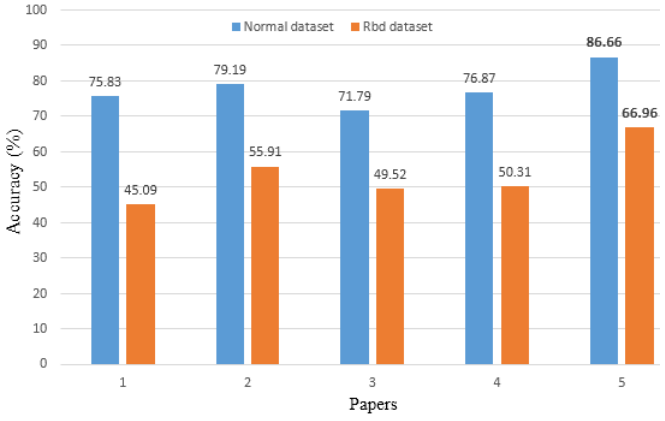

Figure 9: Average classification accuracy over all sleep stages for both normal and *RBD* (patients) datasets for the five papers.

Table 4: Brief description of each method along with their obtained accuracy on normal dataset.

No	Authors	Extracted Features	Feature Selection	Classifier	Result: *mean*(%) (±*std*)					
					Total	W	N1	N2	N3	REM
1	Acharya *et al.* 2010 [75]	4 features; HOS based features	-	GMM Classifier	75.83±0.43	85.67	7.44	75.07	20.69	18.32
2	Fraiwan *et al.* 2010 [87]	21 features; Continuous Wavelet Transform (CWT) using 3 mother wavelets: reverse bioorthogonal (rbio3.3), Daubechies (db20), and Gauss of order 1 (gauss1), Shannon Entropy.	-	LDA+NC Classifier	79.19±0.06	99.12	14.41	89.65	71.42	68.75
3	Gunes *et al.* 2010 [82]	129 features; Welch spectral analysis; k-means clustering based feature weighting (KMCFW).	4 features: Max, Min, Mean, Std of previous features	KNN Classifier	71.79±0.18	89.28	9.54	63.98	38.54	20.90
4	Liang *et al.* 2012 [20]	21 features; Multiscale Entropy (MSE) and Autoregressive (AR) coefficients.	-	LDA+NC Classifier	76.87±0.09	83.90	38.14	51.63	80.77	54.33
5	Fraiwan *et al.* 2012 [88]	7 feature; Continuous wavelet transform (CWT), Choi-Williams distribution (CWD), Hilbert-Huang Transform (HHT) and Renyi's Entropy measures.	-	Random Forest Classifier	86.66±0.09	98.90	18.07	91.11	69.81	71.87

On the other hand, when comparing these five approaches from the robustness point of view, white Gaussian noise is added to the raw signals with different levels of SNR and then the whole process is executed. The SNR levels are 1, 10, 20 dB. Since the results of patients are not comparable to normals, additive noises are just added to the raw normal signals. According to the results which are shown in Fig. 10, The 5^{th} method shows of the lowest sensitivity to noise compared to others, even at different SNR levels.

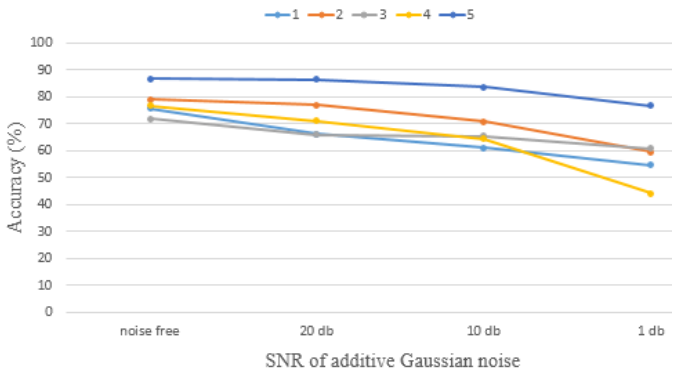

Figure 10: Comparing the accuracy by adding noises with different SNR.

As mentioned earlier, the present study add an additional phase to find the best combination among features and classifiers. As such, the feature sets of each approach is applied to the five classifiers (RF, KNN, LDA+NC, GMM, multi-class SVM). According to the results (Fig 11), random forest produce the best classifier for all different sets of features.

5. Discussion and Conclusion

The aim of this study is to carry out a comprehensive review on automatic sleep stage classification methods. This investigation is performed to give an insight to enlighten the horizon of this field. Sleep scoring is widely used for clinical purposes as well as investigating the brain functions. Although visual sleep scoring is a traditional time consuming method and the accuracy of sleep report highly depends on the expert experience, this method is still a widely-accepted approach. Nevertheless, visual scoring report varies from an expert to another one such that the similarity of two experts on a certain sleep signal do not exceed 83% [7]. Moreover, the limitation of human eyes and brain fatigue make the visual scoring more unstable.

There is a tendency to automatize the sleep scoring process using signal processing methods (Fig. 2). If a simple prediction method is applied to the elicited statistics in Fig. 2, one can expect that this trend will significantly grow in the future. Since the results of quantitative schemes do not suffer from the subjective issues, they can be considered as an efficient auxiliary diagnostic tool which can give valuable information to

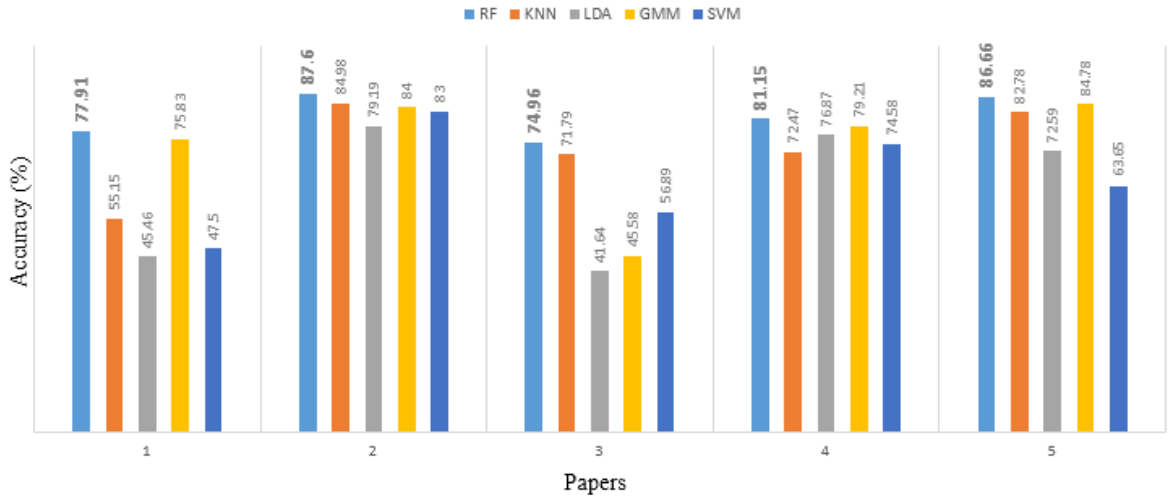

Figure 11: Accuracy of combining different feature sets and classifiers.

sleep specialists. Although several automatic sleep stage classification methods have been suggested [11, 71], a large body of them are applied to EEG signals of healthy subjects while the main concern of automatic methods should be focused on detecting abnormalities in patients' sleep signals.

After reviewing the existing evidences, we conclude that the earlier research has exclusively helped developing feature extraction/selection or modification of a classification method. Among the state-of-the-art studies, five papers are selected and discussed with regard to their performance. The pros and cons of their methods are discussed below.

5.1. Evaluation of Features

The suggested EEG features in those papers (illustrated in Table 4) can be totally divided into four categories: spectrum-base features, time-frequency features, time features, fractal & entropy features.

5.1.1. Spectrum-Base Features

Spectrum estimation methods are repeatedly used in sleep stage classification studies based on which sleep stages are marked with specific frequency ranges and are classified accordingly. To estimate the frequency content of each epoch, parametric (e.g. AR, ARMA), non-parametric (e.g. periodogram, Welch) and HOS methods are vastly utilized. Based on the existing evidences, the elicited features by Welch method could better distinguish the sleep stages than other parametric methods. This supremacy roots in the lower sensitivity of non-parametric methods to te stilled remained motion artifacts and noises as compared to the parametric and cumulant-base methods. This is because non-parametric methods (e.g. Welch method) need to estimate just one auto-correlation function while parametric ones have to construct a matrix of auto-correlation functions. Each auto-correlation function has a bias and variance; therefore, the calculation of auto-correlation matrix while taking an inverse from it, subject to great bias and variance to this estimation. In addition, impact of bias and variance of a cumulant function is much worse than that of an auto-correlation function leading to the deterioration of HOS results.

On the other hand, HOS is capable of revealing the phase-coupling inside of a signal as well as frequency behavior of their corresponding cumulant. Since HOS of a signal presents in the form of a surface in bi-frequency space, it contains unique information and several features can be elicited from this space as compare to other spectrum-base methods. Furthermore, bispectral analysis of EEG signals during the surgical operation can finely estimate the depth of anesthesia [72]. Given the similarity of anesthesia to sleep in some extent, HOS features are expected to provide discriminant values for different stages of sleep. Moreover, higher-order cumulants easily remove the additive Gaussian noise. In the case of sleep signals, if the movement artifact and other additive noises have Gaussian distribution, HOS can eliminate them perfectly. However, cumulant estimation impose a high computational burden, due to several multiplication of the signal with its shifted versions. The main flaw of HOS application raise from te narrow band content of EEGs and also creating a lot of cross-terms in this thin frequency band.

5.1.2. Time Features

Based on the discussed publications, time domain features are frequently used for sleep EEG analysis. The main reason of using such features is that the amplitude of EEG signals significantly varies from one sleep stage to another one. So far, AR coefficients, Hjorth parameters, statistical parameters of the time domain signals (e.g. variance, skewness, kurtosis and MSE), are suggested as discriminative features for sleep analysis. Among the methods discussed, AR model is frequently used for EEG analysis, and also provides good results for discriminating depth of hypnosis [91], identifying the depth of anesthesia [92] as well as sleep stage analysis [20]. As such, AR coefficients can be considered as a versatile feature for different EEG applications. The main advantage of AR model is to encode the time behavior of EEG into few coefficients.

In other word, AR coefficients elicit under-layer dynamics of EEGs through the time domain while other related features may fail to detect such information. Nevertheless, the main flaw of AR model is its high sensitivity to additive noises since these coefficients are determined by auto-correlation functions.

5.1.3. Time-frequency features

Since Fourier transform fails to reveal the frequency and time resolution simultaneously, time-frequency domain transforms are developed. These transforms can be divided in two categories, i.e. linear and nonlinear. Famous linear transforms are the continuous and discrete Wavelet transform and STFT. Nonlinear transforms such as spectrogram and Choi-Williams can demonstrate the energy distribution of a signal, while similar to HOS, they suffer from the effect of cross-terms in a very low frequency band signals like EEGs. Wavelet transform decomposes frequency content of a signal in a dyadic manner matched to the standard EEG frequency bands. In other words, Wavelet acts as a bank of filters to decompose signals into their sub-bands. Since sleep stage transition depends on both the EEG frequency band changes and amplitude variations wavelet features may efficiently detect the stage transition (similar to frequency features). The main advantage of Wavelet features compare to the spectrum-base features is that such features do not involve in calculating inaccurate auto-correlation functions. Moreover, Wavelet transform can be applied to a whole epoch since it is not sensitive to the non-stationary property of EEG signals. In addition, unlike Welch method, no averaging is taken to determine the wavelet coefficients throughout an epoch. In conclusion, Wavelet transform retain comparative advantages to the other time-frequency domain transforms in EEG analysis. It should be noted that for this specific application, CWT provides more separable features than that of DWT in terms of frequency resolution within the same time frame. In addition, wavelet coefficients calculated by CWT are more redundant than DWT. This redundancy provides more stable elicited features for sleep stage classification problem leading to a higher accuracy [11].

5.1.4. Fractal and Entropy Features

Since EEG signals behave irregularly, fractal and entropy methods are suitable to measure the amount of the roughness captured inside the signal. As the entropy of a signal is increased, its information contents is inclined, accordingly. In other words, as a signal behaves more irregular, fractal dimension and entropy of the signal are increased. When a sleep stage is changed, the efficient frequency bandwidth along with the signal amplitude varies accordingly. Therefore, it is expected that fractal and entropy measures can finely describe each sleep stage. Although these measures are somehow correlated to each other, the way of determination of entropy is less sensitive to noise rather than that of the fractal dimension. Accurate computation of fractal dimension and entropy needs a large number of samples making their application limited for analyzing short time signals (e.g. an EEG epoch). Nevertheless, ApEn is a fast entropy measure at the cost of loosing the accuracy. Standard EEG bands are defined by their frequency bandwidth and amplitude variations, while entropy and fractal measures are not designed to capture the signal bandwidth. Therefore, these features might appear effective in detecting just the stage transitions.

5.2. Classifiers

The final phase of automatic sleep stage scoring is classification. Some of the most used classifiers in this field are compared in this study (outlined in Fig. 11). Among the compared classifiers, RF produces the highest sleep stage classification for different types of feature sets introduced in the five selected publications (Fig. 11). The main reason for high efficiency of RF roots into the nature of EEG signals. Since such signals behave irregularly, their elicited features are scattered through the feature space. To distinguish such scattered samples, classic methods including LDA+NC, multi-class SVM, KNN and GMM cannot perform well. As the name implied, RF is designed to classify scattered samples belonging to different classes with a high overlap. For instance, SVM classifier acts upon its support vectors and in a noisy environment, a considerable portion of support vectors become noisy leading to the deterioration of the correct boundary between the two classes.

On the other hand, KNN takes its decision according to the K nearest samples of an input pattern. When the samples of a class are not dense, there is no guaranty whether the neighbors of the given pattern belonging to a certain class. In other words, when the number of samples belong to different classes becomes equal in the vicinity of an input pattern, KNN cannot take any decision.

As far as GMM is a density estimator, each sleep stage is modeled by a certain GMM. When the overlap of the classes is fairly high, their GMM distributions will be similar in terms of mean vectors and covariance matrices. Since the recognition of a sleep stage is performed by finding the maximum value among the GMM outputs, in the case of fairly equal probabilities of different GNMMs (due to the high overlap among the classes), this classifier cannot distinguish the sleep stages precisely.

Further to the above, LDA is a versatile method in EEG signal processing applications [61]. When LDA faces a multi-class dataset, some hyperplanes are constructed to project the input samples into a more separable lower dimension. Since the final decision is made based on NC, the results are not necessary satisfactory. The results from the fact that NC measures the distance of the projected input samples with the center of classes without considering the covariance information of each class. In other words, in real applications, especially when the distribution of a class is multi-modal or skewed, the class-mean is not a good representer of all samples of a given class. Therefore, the decision is failed in practice when the class distribution is not symmetric.

5.3. Natural bottleneck of patients' sleep analysis

Although entropy of CWT features along with RF classifier provide an acceptable performance for healthy subjects, the result of this combination is still unfavorable in patients (Fig. 9).

Since patients with RBD subject to position change and movement, their EEG signals contain high amplitude movement artifacts compared to healthy subjects. Especially in the case of *RBD* patients, their dream is accompanying with muscle contractions leading to changes in the amplitude of REM behaviors and makes it very similar to the wake stage [93]. Thus, regular EEG rhythms become irregular for patients [1] and therefore conventional features brought from other fields (e.g. communications) cannot handle this variation. When analyzing the classification accuracy for each sleep stage separately (shown in Table 4), we see that N1 was repeatedly misclassified due to its high similarity with REM in terms of amplitude and spectrum variations. Moreover, among the sleep stages, N1 has the lowest number of samples causing the employed classifiers fail to learn this class similar to other classes. The classifiers boundary is thus biased toward the classes with higher population. Consequently, a high number of samples in REM stage compared to the N1 one makes the classifier votes in favor of REM. In contrast, due to its difference from other stages (except REM) in terms of frequency content, wake pattern is easily detected. Therefore it ia obvious that this stage is better detected.

6. Future Work

In this research, available studies on sleep stage classification are investigated. As far as the main goal of automatic sleep stage classification methods is detecting abnormal behaviors of EEG signals for the patients, none of the proposed methods achieved a convincing result for patient group. Therefore, developing a customized EEG based features to detect EEG arrhythmia is deemed necessary. Although theses signal processing methods seem to be robust for analyzing artificial and quasi-rhythmic signals (e.g. ECG, radar signal), they are not designed to analyze a chaotic-shape signal like the EEGs of patients with sleep disorders. It should be mentioned that the role of preprocessing and feature extraction parts for this application is more important than the classification part. The most similar sleep stages are N1 and REM which could not be well-separated by the traditional communication based signal processing features. The combinatorial features capturing the roughness, frequency bands and amplitude variations of an EEG epoch can significantly enhance the sleep stage performance.

References

[1] HJ. Park, JS. Oh, DU. Jeong, KS. Park "Automated Sleep Stage Scoring Using Hybrid Rule- and Case-Based Reasoning.", *Computers and Biomedical Research* , 33, pp. 330-349, 2000.

[2] ZT. Yeh, RP. Chiang, SC. Kang, CH. Chiang "Development of the Insomnia Screening Scale based on ICSD-II.", *Int J Psychiatry Clin Pract* , 16(4), pp. 259-267, Oct. 2012.

[3] M. Torabi-Nami, S. Mehrabi, A. Borhani-Haghighi, S. Derman "Withstanding the Obstructive Sleep Apnea Syndrome at the Expense of Arousal Instability, Altered Cerebral Autoregulation and Neurocognitive Decline.", *Journal of Integrative Neuroscience*, 14(2), pp. 169-193, 2015.

[4] M. Younes, W. Thompson, C. Leslie, T. Equan, E. Giannouli "Utility of Technologist Editing of Polysomnography Scoring Performed by a Validated Automatic System.", *Ann Am Thorac Soc*, 12(8) pp. 1206-1218, Aug. 2015.

[5] A. Malhotra, M. Younes, ST. Kuna, R. Benca, CA. Kushida, J. Walsh, A. Hanlon, B. Staley, AI. Pack, GW. Pien, "Performance of an Automated Polysomnography Scoring System Versus Computer-Assisted Manual Scoring.", *Sleep*, 36(4), pp. 573-582, 2013.

[6] NA. Collop, "Scoring variability between polysomnography technologists in different sleep laboratories.", *Sleep Med*, 3(1), pp. 43-50, Jan. 2002.

[7] 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(5), pp. 409-423, May. 2010.

[8] R. Ferri, F. Rundo, L. Novelli, M. G. Terzano, L. Parrino, O. Bruni "A New Quantitative Automatic Method for the Measurement of Non-rapid Eye Movement Sleep Electroencephalographic Amplitude Variability.", *J. Sleep Res*, 21, pp. 212-220, 2012.

[9] C. C. Chiu, B. H. Hai, S. J. Yeh, "Recognition of Sleep Stage Based on a Combined Neural Network and Fuzzy System Using Wavelet Transform Features.", *Biomedical Engineering: Applications, Basis and Communications*, 26(2), pp. 1450021-1450029, Apr. 2014.

[10] S. Motamedi-Fakhr, M. Moshrefi-Torbati, M. Hill ,G. M. Hill, P. r. White, "Signal Processing Techniques Applied to Human Sleep EEG Signals-A Review.", *Biomedical Signal Processing and Control*, 10, pp. 21-33 , 2014.

[11] M. Shokrollahi, K. Krishnan, "A Review of Sleep Disorder Diagnosis by Electromyogram Signal Analysis.", *Critical Reviews^{TM} in Biomedical Engineering*, 43(1), pp. 1-20 , 2015.

[12] VCF. Helland, A. Gapelyuk, A. Suhrbier ,M. Riedl, T. Penzel, J. Kurths "Investigation of an Automatic Sleep Stage Classification by Means of Multiscorer Hypnogram.", *Methods Inf Med.*, 49(5), pp. 467-72 , Jul. 2010.

[13] S. Ozsen, "Classification of sleep stages using class-dependent sequential feature selection and artificial neural network.", *Neural Comput & Applic.*, 23, pp. 1239-1250 , 2013.

[14] ST. Pan, CE. Kuo, JH. Zeng, SF. Liang, "A Transition-constrained Discrete Hidden Markov Model for Automatic Sleep Staging.", *BioMedical Engineering OnLine*, 11, pp. 1-19, 2012.

[15] A. Krakovska, K. Mezeiova, "Automatic Sleep Scoring: A Search for an Optimal Combination of Measures.", *Artificial Intelligence in Medicine*, 53, pp. 25-33, 2011.

[16] ME. Tagluk, N. Sezgin, M. Akin, "Estimation of Sleep Stages by an Artificial Neural Network Employing EEG, EMG and EOG.", *J Med Syst.*, 34, pp. 717-725, 2010.

[17] J. JShi, X. Liu, Y. Li, Q. Zhang, Y. Li, S. Ying "Multi-channel EEG-based Sleep Stage Classification with Joint Collaborative Representation and Multiple Kernel Learning.", *Journal of Neuroscience Methods*, 254, pp. 94-101, 2015.

[18] A. Piryatinska, WA. Woyczynski, MS. Scher, KA. Loparo, "Optimal channel selection for analysis of EEG-sleep patterns of neonates.", *Computer Methods and Programs in Biomedicine* , 106, pp. 14-26, 2012.

[19] S. Khalighi, T. Sousa, G. Pires ,N. Nunes , "Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels.", *Expert Systems with Applications* , 40, pp. 7046-7059 , 2013.

[20] S. 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 Transaction on Instrumentation and Measurement*, 61, pp. 1649-1656, June. 2012.

[21] D. Al-Jumeily , S. Iram , FB. Vialatte, P. Fergus, A. Hussain, "A Novel Method of Early Diagnosis of Alzheimer's Disease Based on EEG Signals.", *Scientific World Journal*, 2015, pp. 1-11, Jan. 2015.

[22] C. Berthomier , X. Drouot , M. Herman-Stoica , P. Berthomier, D. Bokar-Thire, O Benoit, J Mattout, MP d'Ortho "Automatic Analysis of Single-Channel Sleep EEG: Validation in Healthy Individuals.", *Sleep*, 30(11), pp. 1587-1595, Nov. 2007.

[23] T. Penzel , R. Conradt, "Computer Based Sleep Recording and Analysis.", *Sleep Medicine Reviews*, 4(2), pp. 131-148, Nov. 2000.

[24] A. Flexer, G. Gruber, G. Dorffner, "A reliable probabilistic sleep stager based on a single EEG signal.", *Artificial Intelligence in Medicine*, 33, pp. 199-207, 2005.

[25] YL. Hsu, YT. Yang, JS. Wang,CY. Hsu, "Automatic Sleep Stage Recurrent Neural Classifier Using Energy Features of EEG Signals.", *Neuro-*

computing, 104, pp. 105-114, 2013.

[26] DK. Kim, J. Choi, KR. Kim, KG. Hwang, S. Ryu, SH. Cho "Rethinking AASM guideline for split-night polysomnography in Asian patients with obstructive sleep apnea.", *Sleep Breath*, 19(4), pp. 1273-1277, Dec. 2015.

[27] L. Novelli, R. Ferri, O. Bruni, "Sleep classification according to AASM and Rechtschaffen and Kales: effects on sleep scoring parameters of children and adolescents.", *Journal of Sleep Research*, pp. 238-247, Mar. 2010.

[28] A. Rechtschaffen, A. Kales, "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects.", Los Angeles, CA: BIS/BRI, University of California; 1986.

[29] RB. Berry, R. Budhiraja, DJ. Gottlieb, D. Gozal , C. Iber , VK. Kapur, CL. Marcus , R. Mehra, S. Parthasarathy, SF. Quan , S. Redline, KP. Strohl , SL. Davidson Ward , MM. Tangredi "Rules for Scoring Respiratory Events in Sleep: Update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine.", *Sleep Breath*, 19(4), pp. 1273-1277, Dec. 2015.

[30] Y. Kim, S. Kazukawa , M. Kurachi , M. Horita , K. Matsuura, "Problems and a Resolution of Clinical sleep Research-a Recording System of Polysomnography.", *Seishin Shinkeigaku Zasshi*, 93(1), pp. 59-62, 1991.

[31] T. Hori, Y. Sugita , S. Shirakawa, K. Inoue , S. Uchida , H. Kuwahara, M. Kousaka , T. Kobayashi, Y. Tsuji, M. Terashima , K. Fukuda, N. Fukuda,"Proposed Supplements and Amendments to 'A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects', the Rechtschaffen & Kales (1968) standard.", *Psychiatry Clin Neurosci*, 55(3), pp. 305-315, Jun. 2001.

[32] SL. Himanen, j. Hasan, "Limitations of Rechtschaffen and Kales.", *Sleep Med Rev*, 4(2), pp. 149-167, Apr. 2000.

[33] MB. Hamaneh, N. Chitravas, K. Kaiboriboon, S.D. Lhatoo, ; K.A. Loparo, "Automated Removal of EKG Artifact From EEG Data Using Independent Component Analysis and Continuous Wavelet Transformation.", *IEEE Transactions on Biomedical Engineering*, 61(6), pp. 1634-1641, 2014.

[34] M. Crespo-Garcia, M. Atienza, J. L. Cantero, "Muscle Artifact Removal from Human Sleep EEG by Using Independent Component Analysis.", *Annals of Biomedical Engineering*, 36(3), pp. 467-475, Jan. 2008.

[35] C. Stepnowsky, D. Levendowski, D. Popovic, I. Ayappa, DM. Rapoport, "Scoring Accuracy of Automated Sleep Staging from a Bipolar Electrooocular Recording Compared to Manual Scoring by Multiple Raters.", *Sleep Medicine*, 14, pp. 1199-1207, 2013.

[36] J. Muthuswamy, N. V. Thakor "Spectral analysis methods for neurological signals.", *Journal of Neuroscience Methods*, 83(1), pp. 1-14, Aug. 1998.

[37] B. Weiss, Z. Clemens, R. Bodize, Z. Vago, P. Halasz, "Spatio-temporal Analysis of Monofractal and Multifractal Properties of the Human Sleep EEG.", *Journal of Neuroscience Methods*, 185(1), pp. 116-124, Dec. 2009.

[38] B. Hjorth, "EEG Analysis Based on Time Domain Properties.", *Electroencephalography and Clinical Neurophysiology*, 29(3), pp. 306-310, Sept. 1970.

[39] B. Koley, D. Dey, "An ensemble system for automatic sleep stage classification using single channel EEG signal.", *Comput Biol Med.*, 42(12), pp. 1186-1195, Oct. 2012.

[40] LJ. Herrera, CM. Fernandes, AM. Mora, D. Migotina, R. Largo, A. Guillen, AC. Rosa, "Combination of Heterogeneous EEG Feature Extraction Methods and Stacked Sequential Learning for Sleep Stage Classification.", *Int J Neural Syst.*, 23(3), pp. 1-20, 2013.

[41] GA. Correa, E. Laciar, HD. Patino, ME. Valentinuzzi, "Artifact removal from EEG Signals Using Adaptive Filters in Cascade.", *J Phys Conf Ser.*, 90, pp. 1-10, 2007.

[42] RJ. Croft, JS. Chandler, RJ. Barry , NR. Cooper, AR. Clarke "EOG correction: a comparison of four methods.", *Psychophysiology*, 42(1), pp. 16-24, Jan. 2005.

[43] C.F.V. Latchoumane, J. Jeong, "Quantification of brain macrostates using dynamical nonstationarity of physiological time series.", *IEEE Trans. Biomed.Eng*, 58(4), pp. 1084-1093, Apr. 2011.

[44] F. Karimzadeh, E. Seraj, R. Boostani , M. Torabi-Nami, "Presenting Efficient Features for Automatic CAP Detection in Sleep EEG Signals.", *38th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 448-452, July. 2015.

[45] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, F. Chapotot "Feature Selection for Sleep/wake Stages Classification Using Data Driven Methods.", *Biomedical Signal Processing and Control*, 2, pp. 171-179, 2007.

[46] C. Vural, M. Yildiz, "Determination of Sleep Stage Separation Ability of Features Extracted from EEG Signals Using Principle Component Analysis.", *J Med Syst.*, 34, pp. 83-89, 2010.

[47] JA. Jiang, CF. Chao,MJ. Chiu , RG. Lee, CL. Tseng. R. Ling "An Automatic Analysis Method for Detecting and Eliminating ECG Artifacts in EEG.", *Computers in Biology and Medicine*, 37(11), pp. 1660-1671, Nov. 2007.

[48] C. M. Bishop, "Pattern Recognition and Machine Learning", *Springer-Verlag New York*, 1st Edition, 2006.

[49] T. Kayikcioglu, M. Maleki, K. Eroglu "Fast and Accurate PLS-based Classification of EEG Sleep Using Single Channel Data.", *Expert Systems with Applications*, 42, pp. 7825-7830, 2015.

[50] H. Akaike, "New look at statistical-model identification", *IEEE Trans. Automatic Control AC19*, pp. 716-723, 1974.

[51] L Sörnmo, P Laguna, "Bioelectrical signal processing in cardiac and neurological applications.", *USA: Elsevier Academic Press*, 2005.

[52] K. Susmakova, A. Krakovska, "Discrimination Ability of Individual Measures Used in Sleep Stages Classification.", *Artificial Intelligence in Medicine*, 44, pp. 261-277, 2008.

[53] R. Esteller, G. Vachtsevanos, J. Echauz and B. Litt, "A comparison of Waveform Fractal Dimension Algorithms", *IEEE transactions on circuits and sytems* , 48 (12), pp. 177-183, Feb. 2001.

[54] B. Sen, M. Peker, A. Cavusoglu, F. V. Celebi, "A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms.", *Journal of Medical Systems*, 28(3), pp. 1-21, Mar. 2014.

[55] BL. Su, Y. Luo, CY. Hong, ML. Nagurka, CW. Yen "Detecting slow wave sleep using a single EEG signal channel.", *Journal of Neuroscience Methods*, 243, pp. 47-52, 2015.

[56] RU. Acharya, O. Faust, N. Kannathal, T. Chua "Non-linear Analysis of EEG Signals at Various Sleep Stages.", *Computer Methods and Programs in Biomedicine* , 80, pp. 37-45, 2005.

[57] A. Piryatinska, G. Terdik, WA. Woyczynski, KA. Loparo, MS. Scher, A. Zlotnik "Automated detection of neonate EEG sleep stages.", *Computer Methods and Programs in Biomedicine* , 95, pp. 31-46, 2009.

[58] A. Brignol, T. Al-ani, X. Drouot, "Phase Space and Power Spectral Approaches for EEG-based Automatic Sleep-wake Classification in Humans: A Comparative Study Using Short and Standard Epoch Lengths.", *Computer Methods and Programs in Biomedicine*, 109(3), pp. 227-238, 2013.

[59] C. Cortes, V. Vapnik, "Support-vector networks.", *Machine Learning*, 20(3), pp. 273-297, Sept. 1995.

[60] HT. Wu, R. Talmon, YL. Lo, "Assess Sleep Stage by Modern Signal Processing Techniques.", *IEEE Transactions on Biomedical Engineering*, 62(4), pp. 1159-1168, Dec. 2012.

[61] M. Sabeti, S. Katebi, R. Boostani, "Entropy and complexity measures for EEG signal classification of schizophrenic and control participants.", *Artificial Intelligence in Medicine*, 243, pp. 263-274, 2009.

[62] B. Xue, M. Zhang, and WN. Browne. "Particle Swarm Optimization for Feature Selection in Classification: Novel Initialization and Updating Mechanisms.", *Applied Soft Computing* 18, pp. 261-276, 2014.

[63] RO. Duda, PE. Hart, DG. Stork, "Pattern classification.", 2nd ed., New York, *John Wiley & Sons Ltd*., 2001.

[64] L. Breiman, "Random forests", *Machine Learning*, 45 (1), pp. 5-32, 2001.

[65] M. Ronzhina, O. Janousek, J. Kolarova, M. Novakova, P. Honzik "Sleep Scoring Using Artificial Neural Networks.", *Sleep Medicine Reviews*, 16, pp. 251-263, 2012.

[66] RK. Sinha, "Artificial Neural Network and Wavelet Based Automated Detection of Sleep Spindles, REM Sleep and Wake States.", *J Med Syst.*, 32, pp. 291-299, 2008.

[67] J. MacQueen, "Some methods for classification and analysis of multivariate observations.", Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, pp. 281-297, University of California Press, Berkeley, Calif., 1967.

[68] R. Agarwal, J. Gotman, "Computer-Assisted Sleep Staging.", *IEEE Transactions on Biomedical Engineering*, 48(12), pp. 1412-1423, Dec. 2001.

[69] P. Berkhin, "A Survey of Clustering Data Mining Techniques."*Grouping*

multidimensional data. Springer Berlin Heidelberg, 25-71, 2006.

[70] F. Murtagh, "A Survey of Recent Advances in Hierarchical Clustering Algorithms.", *The Computer Journal* 26(4), pp. 354-359, 1983.

[71] RA. Jarvis, EA. Patrick "Clustering Using a Similarity Measure Based on Shared Near Neighbors.", *IEEE Transactions on Computers* 22(11), pp. 1025-1034, 1973.

[72] KC. Chua, V. Chandran, R. Acharya, CM. Lim, "Application of Higher Order Statistics/spectra in Biomedical Signals-A Review.", *Medical Engineering & Physics*, 32(7), pp. 679-689, Sept. 2010.

[73] B. Mandelbrot, "How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension.", Science 156 (3775), pp. 636-638, 1967.

[74] C. Fraley, and AE. Raftery. "Model-based clustering, discriminant analysis, and density estimation."*Journal of the American statistical Association* 97, pp. 611-631.,2002

[75] U. Acharya, EC. Chua, KC. Chua, LC. Min, T. Tamura "Analysis and Automatic Identification of Sleep Stages Using Higher Order Spectra.", *International Journal of Neural Systems*, 20(6), pp. 509-521, 2010.

[76] DA. Reynolds, RC. Rose "Robust Text-Independent Speaker Identification Using Gaussian Mixture Model.", *IEEE transaction on speech and audio processing*, 3(1), pp. 72-83, Jan 1995.

[77] F. Yaghouby, S. Sunderam, "Quasi-supervised scoring of human sleep in polysomnograms using augmented input variables.", *Computers in Biology and Medicine*, 59, pp. 54-63, 2015.

[78] A. P. Petropulu, "Higher-Order Spectral Analysis."*The Biomedical Engineering Handbook*: Second Edition. Ed. Joseph D. Bronzino Boca Raton: CRC Press LLC, 2000

[79] T. Schluter, S. Conrad, "An Approach for Automatic Sleep Stage Scoring and Apnea-hypopnea Detection.", *Front. Comput. Sci.*, 6(2), pp. 230-241, 2012.

[80] V. Bajaj, RB. Pachori, "Automatic Classification of Sleep Stages based on the Time-frequency Image of EEG Signals.", *Computer Methods and Programs in Biomedicine*, 112(3), pp. 320-328, Dec. 2013.

[81] SF. Liang, CE. Kuo, YH. Hu, YS. Cheng, "A Rule-based Automatic Sleep Staging Method.", *Journal of Neuroscience Methods*, 205, pp. 169-176, 2012.

[82] S. Gunes, K. Polat, S. Yosunkaya, "Efficient Sleep Stage Recognition System Based on EEG Signal Using K-means Clustering Based Feature Weighting.", *Expert Systems with Applications*, 37, pp. 7922-7928, 2010.

[83] K. Fukunaga, "Introduction to Statistical Pattern Recognition.", *Academic press*, , 2nd Edition, 2013.

[84] AL. Goldberger, L. Amaral, L. Glass, JM. Hausdorff, PCH. Ivanov, RG . Mark, JE. Mietus, GB. Moody , CK. Peng , HE. Stanley . PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):215-220.

[85] B. Kemp, AH. Zwinderman, B. Tuk, HAC. Kamphuisen, JJL. Oberye "Analysis of a Sleep-dependent Neuronal Feedback Loop: The Slow-wave Microcontinuity of the EEG.", *IEEE-BME*, 47(9), pp. 1185-1194, 2000.

[86] MG. Terzano, L. Parrino, A. Sherieri, R. Chervin, S. Chokroverty, C. Guilleminault, M. Hirshkowitz, M. Mahowald, H. Moldofsky, A. Rosa, R. Thomas, A. Walters "Atlas, Rules, and Recording Techniques for the Scoring of Cyclic Alternating Pattern (CAP) in human sleep.", *Sleep Med.*, 2(6), pp. 537-553, Nov. 2001.

[87] 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.*, 43(3), pp. 230-237, 2010.

[88] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, H. Dickhaus, "Automated Sleep Stage Identification System Based on Timefrequency Analysis of a Single EEG Channel and Random Forest Classifier.", *Computer Methods and Programs in Biomedicine*, 108, pp. 10-19, 2012.

[89] T. Lajnef, S. Chaibi, P. Ruby, PE. Aguera, JB. Eichenlaub, M. Samet, A. Kachouri, K. Jerbi, "Learning Machines and Sleeping Brains: Automatic Sleep Stage Classification Using Decision-tree Multi-class Support Vector Machines.", *Journal of Neuroscience Methods*, 250, pp. 94-105, Jul. 2015.

[90] L. Fraiwan, K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, H. Dickhaus, "Time Frequency Analysis for Automated Sleep Stage Identification in Fullterm and Preterm Neonates.", *J Med Syst.*, 35, pp. 693-702, 2011.

[91] Z. Elahi, R. Boostani, and A. Motie Nasrabadi. "Estimation of hypnosis susceptibility based on electroencephalogram signal features.", *Scientia Iranica* 20(3), pp. 730-737, 2013.

[92] EW. Jensen, P. Lindholm, and S. W. Henneberg, "Autoregressive modeling with exogenous input of middle-latency auditory-evoked potentials to measure rapid changes in depth of anesthesia.", *Methods of information in medicine*, 35(3), pp. 256-260, 1996.

[93] ML Fantini, JF Gagnon, D Petit, S Rompre, A Decary,J Carrier, J Montplaisir. "Slowing of electroencephalogram in rapid eye movement sleep behavior disorder.", *Annals of neurology*, 53(6), pp. 774-780, 2003.