This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2016.2552539, IEEE Transactions on Neural Systems and Rehabilitation Engineering

EEG Sleep Stages Classification Based on Time Domain Features and Structural Graph Similarity

Mohammed Diykh, Yan Li and Peng Wen

School of Agricultural, Computational and Environmental Sciences
University of Southern Queensland, Australia
{Mohammed.Diykh; Yan.Li; Peng.Wen}@usq.edu.au

***Abstract*-The electroencephalogram (EEG) signals are commonly used in diagnosing and treating sleep disorders. Many existing methods for sleep stages classification mainly depend on the analysis of EEG signals in time or frequency domain to obtain a high classification accuracy. In this paper, the statistical features in time domain, the structural graph similarity and the K-means (SGSKM) are combined to identify six sleep stages using single channel EEG signals. Firstly, each EEG segment is partitioned into sub-segments. The size of a sub-segment is determined empirically. Secondly, statistical features are extracted, sorted into different sets of features and forwarded to the SGSKM to classify EEG sleep stages. We have also investigated the relationships between sleep stages and the time domain features of the EEG data used in this paper. The experimental results show that the proposed method yields better classification results than other four existing methods and the support vector machine (SVM) classifier. A 95.93% average classification accuracy is achieved by using the proposed method**.

Index Terms: **EEG signal, Structural graph similarity, Time domain features, Sleep stages.**

I. INTRODUCTION

EEG signals are measured using electrodes placed on the scalp. They record the electrical potentials generated by the nerve cells in the brain and open a window for the investigation of neural activities and brain functioning [1, 19]. EEG signals are an important source for studying human brain activities and for diagnosing and monitoring neurological diseases, such as sleep disorders and epilepsy.

We spend about one third of our life asleep. Human sleep is a dynamic process which can be divided into two main states: the rapid eye movement (REM) and the non-rapid eye movement (NREM), whereas the latter can be further divided into 4 stages, namely, Stages 1, 2, 3 and 4 [1, 30]. Each sequential stage of the NREM sleep stages is indicative of a deeper sleep, with Stage 1 being the lightest and Stage 4 the deepest [25]. A typical night's sleep consists of 75% NREM sleep. The REM is a status stage where dreams occur and constitutes 25% of a normal sleep night [21, 25].

When used for diagnosis, physicians are in favour of EEGs instead of other biological signals, such as electromyography (EMGs) and electrooculography (EOGs) [7, 9, 19]. As a result, many commonly used medical classification systems are developed based on EEG signals [5, 19, 26, 39, 43]. In recent years, much research work has been conducted in EEG classification. Usually, such classification is completed using single channel EEG signals [15, 43]. Those work focused on classifying sleep stages using different techniques, such as neural networks [19], support vector machines (SVM) [36, 43] and k-means [11, 26]. But it is often difficult to achieve a high level of accuracy [34].

Graph theory is increasingly used in neuroscience research to study brain diseases and healthy subjects [27, 39]. It is a powerful tool for characterizing the functional topological properties of brain networks involved both in normal and abnormal brain functioning [4, 12, 20, 32, 35, 38]. In addition, graph theory has been proved to be very useful in statistics to describe the relations among random variables [27, 31, 35]. In recent years, it has been widely used for classifying and analysing the relationships in complex networks, such as biological and brain networks, social networks, signal and image processing. In image processing, graph theory is considered as one of the most powerful tools to classify and analyse digital images. In this work, graph theory is employed to classify sleep stages using EEG signals.

In this paper, the time domain features and structural graph similarity combined with K-means (SGSKM) are used to achieve a high level of classification accuracy, which is much better than the manual results obtained by the experts (about 83%). In this work, each segment from an EEG signal is partitioned to *m* sub-segments to make the signal stationary. The number of sub-segments (*m*) is empirically determined by experiments. Then the statistical features are extracted from each sub-segment to reduce the data dimension. A vector of features from each segment is extracted and then mapped into a graph. Finally, the structural graph properties are extracted from each of the graphs to represent the original EEG signal. The K-means algorithm is applied to classify the graph representation features into six sleep stages groups, with each group representing one sleep stage.

In order to evaluate the performance of the proposed method, the comparisons are made among the proposed method and four other existing methods as well as the SVM. The experimental results show that the proposed method gives better accuracy results (95.93%), compared with other methods. This technique has the potential to improve the existing EEG sleep stages classification techniques and help physicians with auto sleep stages scoring.

The rest of the paper is structured as follows: Section II presents a brief overview about the current EEG sleep stages classification techniques. In Section III, the EEG datasets used in the paper are described. Section IV introduces the proposed SGSKM method and discusses the performance evaluation measures. Section V presents the experimental results and compares the results from the proposed method and those from other popular methods. Finally, Section VI draws the conclusions from this study.

II. RELATED WORK

In this section we review some of the related research work. In most research, the representative features are extracted from EEG signals, and then these features are forwarded to a classifier to distinguish the EEG signal into six or seven sleep stages.

In the literature, the wavelet transform has been widely used for sleep EEG stages analysis [6, 11, 24, 25, 42, 44] Oboyya et al. [25] classified six sleep stages using a single channel EEG signal. The wavelet transform was utilized to extract the main features from EEG signals. Then a matrix of the main features was forwarded to a fuzzy C-means algorithm to classify the six sleep stages. The average of the classification rate reported was 85%. However, the ages of the subjects used in that work were limited between 35 and 50.

Güneş et al. [11] identified sleep stages using the K-means clustering method based on feature weighting. Welch spectral transform was used to extract the features from EEG signals. 196 features were obtained in their study. Some statistical features were used to reduce the number of the obtained features. Finally the K-means and decision tree techniques were used to classify the sleep states into six stages. The accuracy percentage was scored around 83%. The study was conducted with just five male subjects.

Naderi at el. [24] also employed Welch spectrum analysis and a neural network to classify sleep stages. The Welch spectrum analysis was used to decompose and extract the features from the EEG signals. The extracted features were forwarded to a recurrent neural network. The data used in that research was from the Sleep-EDF (Europe data format) database.

Zhovna and Shallom [42] used a multi-channel EEG signal modelling method to classify sleep stages. In the first part, a multichannel auto regressive model was used to extract the main parameters from the EEG signals. In the second part, the classification process was achieved using Kullback-Leibler divergence method. The classification accuracy percentage obtained in the study was 90%.

More recently, Zhu et al. [43] utilized the concepts of visibility graphs and horizontal visibility graphs to extract the features from EEG signals. The representative graph features were extracted, and then were forwarded to a support vector machine to classify the six sleep stages. The datasets used in the research were from the Sleep-EDF database. It was reported that an average of 87% classification accuracy was achieved.

Suily et al. [33] developed and applied a new technique that was called clustering technique based least square support vector machine to classify EEG signals. The classification method was conducted in two phases. In the first phase, the clustering technique was used to extract the features. Then, the least square support vector machine was utilized to classify the extracted features. The proposed method was conducted with three publicly available datasets: epileptic EEG, motor imagery EEG data and mental imagery tasks EEG. The average classification accuracy for the three datasets were 94.55%, 84.52% and 61.60%, respectively.

Bajaj and Pachori [2] classified the sleep stages based on the time frequency images (TFIs) of EEG signals by utilizing a smooth pseudo Wigner-Ville distribution . The frequency bands of rhythms of EEG signals were used to segment the TFIs. The histograms of the segmented TFIs were forwarded to a multiclass least squares support vector machine for automatically classifying the sleep stages.

Aboalayon et al. [1] also classified the sleep stages based on EEG signals. Butterworth band pass filters were used to decompose the EEG signals into five bands (delta, theta, alpha, beta and gamma). Different features were extracted from the five bands, and these features were fed to a SVM to classify them into six sleep stages. The dataset used in the study was publicly available online from the Sleep-EDF database. The same datasets were used as the ones used in this paper. The study reported a 90% classification accuracy.

Hsu et al. [13] introduced a recurrent neural classifier for sleep stages classification. The energy features were extracted from the characteristics of an EEG signal (fpz-cz channel). The recurrent neural classifier distinguished the energy features extracted from each epoch (30 seconds of an EEG signal) into one of the sleep stages. Eight sleep recording files were used in the work, which were obtained from the sleep-EDF database.

Vural and Yildiz [39] made a study to determine the ability of identifying sleep stages using different feature sets extracted from sleep EEG signals. In this study three types of features were used to represent the original EEG signals. They were time domain, frequency domain (spectral power analysis) and hybrid features which included both the time domain and frequency domain features. The principle component analysis was used to determine the best feature set that could determine the EEG sleep stages correctly.

III. EXPERIMENTAL DATA

In this work, two publicly available datasets are used for the proposed method to classify EEG sleep stages. The following section gives brief explanations of the two datasets.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2016.2552539, IEEE Transactions on Neural Systems and Rehabilitation Engineering

A. Sleep-EDF database (Dataset-1)

The dataset was obtained from the Sleep-EDF data [17, 10]. The dataset is available for the public to access[1]. In the dataset, there were eight sets of EEG data collected and all were selected for the experiments in this paper. The datasets were reordered from different volunteers in 1989 and 1994, respectively, and stored in the EDF format. The demographics information of subjects were: The first four records (sc4112e0 to sc4002e0) were recorded in 1989 from Caucasian males and females healthy volunteers during 24 hours in their normal daily life, aged from 26-101. The other four subjects named (st7132j0 to st7022j0) were recorded in 1994 also from Caucasian males and female who had mild difficulty falling asleep but were otherwise healthy. Their age was from 18-34 years old [18, 23]. Each file contains one horizontal channel, two EEG signal channels (Fpz-Cz and Pz-Oz), Resporonasal, EMGSubmenta, Tempbody, and Eventmarker.

In this study we used channel Pz-Oz because it provided a better classification performance than Fpz-Cz channel [29, 43]. The original sleep stages of these segments were labelled as AWA, S1, S2, S3, S4, REM MVT (movement time), and UNS (unknown state). Each whole EEG signal was divided into segments in 30 seconds (containing 3000 points of data). The recordings were also scored in sleep stages according to the R& K criteria [28] with every 30 seconds of an EEG signal. Table 1 shows the number of segments that are used in this study.

TABLE 1
INFORMATION OF THE EXPERIMENTAL DATA (DATASET-1)

Sleep stage	No. of Segments
AWA	7830
S1	603
S2	3621
S3	672
S4	627
REM	1609
Total no. of segments	14963

B. Sleep Spindles database (Dataset-2)

The EEG sleep stages database which is publically available online[2] is also used in this study. They were recorded at the University of MONS-TCTS laboratory and the University of Libre de Bruxelles - CHU de Charleroi Sleep Laboratory [37]. The recordings were obtained from eight males and females (aged from 31-54). They consist of 20 whole night recordings. Eight recordings were used in this study. The recordings contain two EOG channels (P8-A1, P18-A1), three EEG channels (CZ-A1 or C3-A1, FP1-A1 and O1-A1) and one submental EMG channel. In this study only C3-A1 channel was used. The data were sampled at frequency of 200 Hz. The EDF standard was utilized for storing. The recordings were also scored in sleep stages according to the R&K criteria [28]. Table 2 shows the number of segments that are used in this study.

TABLE 2
INFORMATION OF THE EXPERIMENTAL DATA (DATASET-2)

Sleep Stage	No. of segments
AWA	8858
S1	2856
S2	21184
S3	3216
S4	4441
REM	8000
Total no. of segments	48555

IV. THE METHODOLOGY

This paper proposes an efficient method for classifying EEG signals into six sleep stages. Fig. 1 illustrates the structure of the proposed method. Each segment of the EEG signals is divided into *m* sub-segments. The statistical features from each sub-segment are extracted. The vector of the features from each segment is then mapped into a graph. For each graph the structural similarity properties are extracted. The graph features are then forwarded to a classifier to classify each segment into one of the six sleep stages. The details of the methodology are explained in the following sections. We also apply the SVM to the two datasets for the comparison with the proposed method.

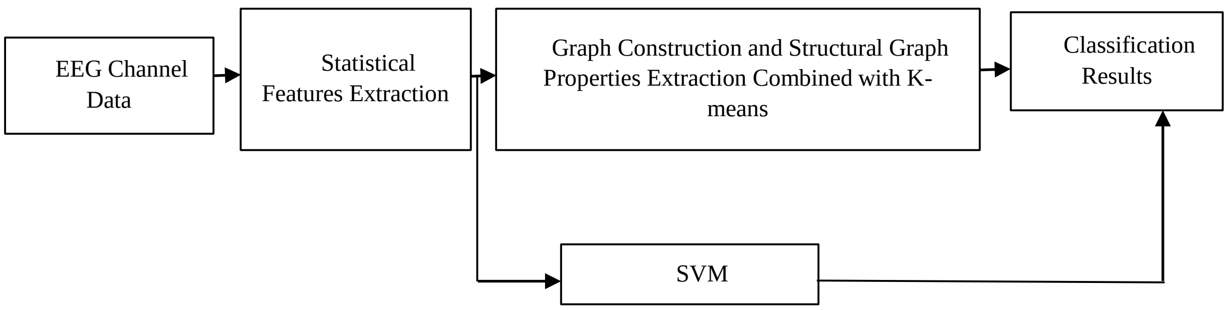


Fig. 1. Diagram of the proposed method

[1] http://www.physionet.org/physiobank/database/sleep-edf/

[2] http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseSubjects

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2016.2552539, IEEE Transactions on Neural Systems and Rehabilitation Engineering

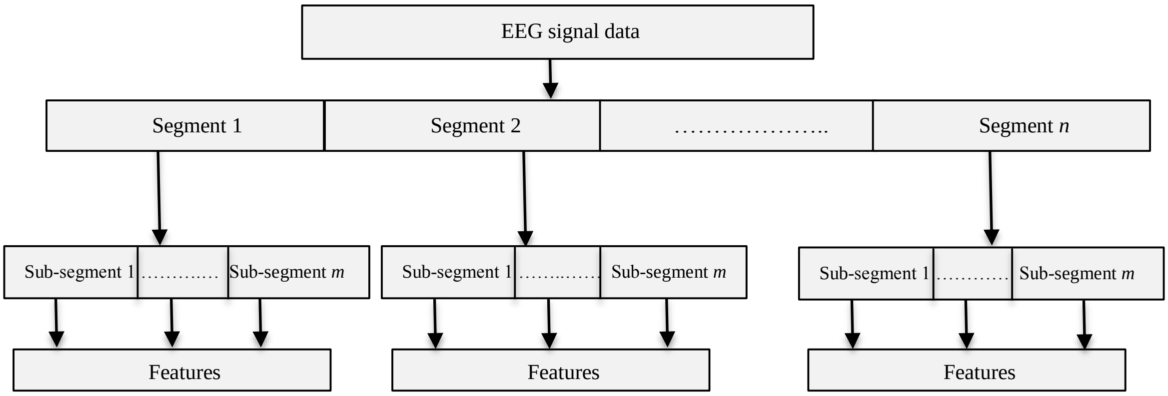


Fig.2. Segmentation and statistical features extraction

A. Statistical Features Extraction

Feature extraction is the most important part of the classification because if the features are not chosen well the performance of a classification will be degraded. It is even more important to design an effective method to extract features from EEG signals, because they are also periodic. One effective method to quantifying a non-stationary time series such as EEG signals is to view it as consisting of many number of segments that are themselves stationary. In this work, the EEG signals are divided into smaller segments. The interval of each segment is 30 seconds (3000 data points) in this paper according to the Rechtschaffen and Kales [28]. Then, each segment is further divided into *m* sub-segments for a shorter period of time to make the signals stationary and $m \geq 1$. The number of sub-segments from each segment and the length of each sub-segment are empirically determined during the experiments. The optimum number of sub-segments is determined according to the following algorithm.

Input *X= EEG signal*

Algorithm

1. *let p be a number that each EEG segment should be divided, initially p=0*
2. *k= first segment of X*
3. ***For each** EEG segment **do***
4. *p=p+1*
5. ***divide** k into p sub-segments*
6. ***extracting** the statistical features from each sub-segment*
7. ***forward** the extracted features to the proposed method*
8. ***if** the classification accuracy is satisfactory then stop segmentation , got to step 10*
9. *Else go to step 4*
10. *End*

Output *Set of statistical features*
Number of sub-segments

As the result, each segment is separated into 75 sub-segments, with each sub-segment includes 40 data points each. Fig. 2 shows the segmentation and features extraction technique. In the experiments, we use a statistical approach to extract the time domain features from each sub-segment, and then put all the features from one segment in a vector to represent the segment. It is found that some of EEG data are symmetric distribution and other skewed distribution. The mean and the standard division are considered appropriate measures for a time series with symmetric distribution. While for a skewed distribution median, range and quartile are effective to measure the centre and the spread of a dataset. However, feature mode which defines as the most frequent value is used to measure the locations of a time series. Other statistical features, such as minimum, variation, skewness and kurtosis are used as measures to pull out the important information about a time series. For these reasons, 12 features of *{median, maximum, minimum, mean, mode, range, first quartile, second quartile, standard deviation, variation, skewness, kurtosis}* are considered as key features to represent EEG data in this study. Suily et al. [33] used statistical properties as the key features to classify motor images and epileptic data. Sen et al. [34] utilized statistical features with time frequency features in a comparative study in EEG classification. Table 3 provides a short explanation of the features. The 12 features are denoted as $\{X_{Me}, X_{Max}, X_{Min}, X_{Mean}, X_{Mod}, X_{Rand}, X_{Q1}, X_{Q2}, X_{SD}, X_{Var}, X_{Ske}, X_{Kue}\}$. These segment features are then mapped into a graph. The structural graph similarity of all the graphs is used to classify EEG segments into different sleep stages. In Section V, several experiments are designed to test which combination of the features is the best to represent the EEG data.

B. Complex Network and Structural Graph Similarity

Graph theory provides many global and local quantitative measures to analyse the brain network dynamics. In this work, we calculated a topological graph structure property to classify six sleep stages in an EEG signal. Statistical analysis was conducted to extract the main features from graph nodes. According to Zhang and Small [41], each vector of features (one segment) was mapped as an undirected graph **G**= (**V**, **E**), where **V** denotes the set of nodes and **E** the set of connection among the nodes. The connection between each pair of nodes refers to the existence of a relationships between the nodes [3, 12, 22]. One of the commonly used similarity measuring method for a time series is the Euclidean distance [4, 14, 16]. Let Let $\{x_{ij}\} = 1, 2, 3\ldots N$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2016.2552539, IEEE Transactions on Neural Systems and Rehabilitation Engineering

TABLE 3

SHORT EXPLANATION OF THE STATISTICAL FEATURES

No.	Feature name	Formula	No.	Feature name	Formula
1	Maximum	X_{Max}=Max[x_n]	7	Minimum	X_{Min}=min[x_n]
2	Mean	$X_{Mean} = \frac{1}{n}\sum_{1}^{n} x_i$	8	Mode	$X_{Mod} = L + \left(\frac{f_1 - f_0}{2f_1 - f_2}\right)Xh$
3	Median	$X_{Me} = (\frac{N+1}{2})^{th}$	9	Range	$X_{Rang}=X_{Max}$-X_{Min}
4	First Quartile	$X_{Q1} = \frac{1}{4(N+1)}$	10	Standard Deviation	$X_{SD} = \sqrt{\sum_{n=1}^{N}(x_n - AM)}\,\frac{2}{n-1}$
5	variation	$X_{Var} = \sum_{n=1}^{N}(x_n - AM)\frac{2}{N-1}$	11	Skewness	$X_{Ske} = \sum_{n=1}^{N}(x_n - AM)\frac{3}{(N-1)SD^3}$
6	Kurtosis	$X_{Ku} = \sum_{n=1}^{N}(x_n - AM)\frac{4}{(N-1)SD^4}$	12	Second Quartile	$X_{Q2} = \frac{4}{4(N+1)}$

where X_n=1, 2, 3, …, *n*, is a time series, *N* is the number of data points, AM is the mean of the sample.

be a set of time series of *N* data points. Each data point of the series was assigned to be a node in an undirected graph. Two nodes, v_1 and v_2, in a graph are connected if a distance between the two nodes are less than or equal to an adaptive threshold.

$$(v_1, v_2) \in E, if\ d(v_1, v_2) \leq \delta \quad (1)$$

where δ is an adaptive threshold, Fig. 3 shows a time series: {2.4, 5.0, 6.2, 6.0, 7.5, 8.5, 3.0, 1.8, 9.5} being transferred into a graph. The adjacency matrix **A** of graph **G** is calculated for all **V** to describe the graph nodes connected. The adjacent matrix of an undirected graph is symmetric, i.e. A (v_i, v_j)= A (v_j, v_i)

$$\boldsymbol{A}(v_i, v_j) = \begin{cases} 1, \text{if } (v_i, v_j) \in \boldsymbol{E} \\ 0, \quad \text{other wise} \end{cases} \quad (2)$$

The number of nodes increases when there are more sub-segments in one EEG segment, and vice versa. All graphs have a fixed number of nodes in this paper. From Fig. 3 we can notice that, node v_8 has no connections with other nodes in the graph. This means that node v_8 is an isolated point in the graph.

C. Graph Features

The statistical features of a graph can be obtained from the adjacency matrix of the graph [20, 8, 40].

The following features are extracted and used for classification in this study.

• Degree Distribution

The degree distribution for a graph, denoted by $\mathbf{P}(k)$, is defined to be the fraction of nodes in the graph with a degree k. The degree distribution can be calculated as:

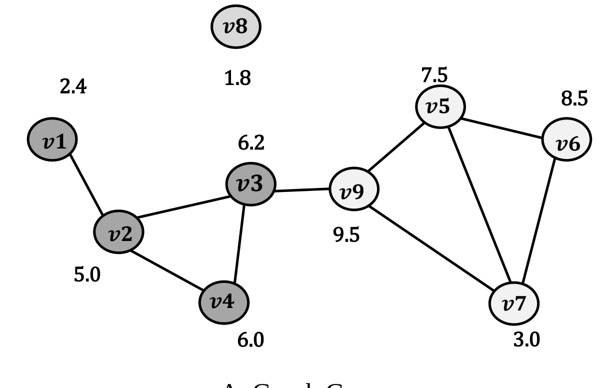
$$\mathbf{P}(k) = \frac{|\{v|d(v) = k\}|}{N} \quad (3)$$

where $d(v)$ is the degree of node v, and N is the number of nodes in the graph.

• *Clustering Coefficient*

Clustering coefficient (CC) is frequently used to characterize the global and local structures of a graph [20, 22, 35]. Stam et .al [35] and Li et.al [20] used the CC to analyse brain activities. The clustering coefficient for node v_i in graph **G** is defined as $C_{v_i} = \frac{L}{\frac{N_G}{2}}$, where L is the number of the actual links between v_i with its neighbours, and N_G is the number of the neighbours of v_i.
The clustering coefficient of the graph is the average of the clustering coefficients of all the nodes.

$$C = \frac{1}{N}\sum_{i=1}^{N} C_{v_i} \quad (4)$$



A. Graph G

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
v_1	0	1	0	0	0	0	0	0	0
v_2	1	0	1	1	0	0	0	0	0
v_3	0	1	0	1	0	0	0	0	1
v_4	0	1	1	0	0	0	0	0	0
v_5	0	0	0	0	0	1	1	0	1
v_6	0	0	0	0	1	0	1	0	0
v_7	0	0	0	0	1	1	0	0	1
v_8	0	0	0	0	0	0	0	0	0
v_9	0	0	1	0	1	0	1	0	0

B. Adjacency matrix of graph G

Fig. 3. Example of a time series converted to graph

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2016.2552539, IEEE Transactions on Neural Systems and Rehabilitation Engineering

where N is the number of nodes in the graph and C_{v_i} is the CC for node v_i. To calculate the CC of node v_3 in Fig. 3, the three neighbours of v_3 are connected each other ($v_4 \rightarrow v_9$, $v_3 \rightarrow v_9$, $v_4 \rightarrow v_3$). Thus, CC (v_3)= 1.

- *Jaccard Similarity Coefficients*

Jaccard coefficient is a statistic function used for comparing the similarity and the diversity of two nodes in a graph [12]. It is evaluated as the set of the intersection neighbours between two nodes divided by the neighbour set of the union of the two nodes. Jaccard coefficient function is calculated based on the following equation:

$$\boldsymbol{w}(v_1, v_2) = \frac{|\Gamma(v_i) \cap \Gamma(v_j)|}{|\Gamma(v_i) \cup \Gamma(v_j)|} \qquad (5)$$

where Γ (v_i) is the set of neighbors of node v_j, Γ (v_j) is the set of neighbors of v_j, and $\boldsymbol{w}$ = [0, 1]. For each graph, a Jaccard coefficient vector is computed.

D. Classification

After the structural graph properties are obtained from each graph, we classify the segment based on the structural graph features. The K-means clustering algorithm [11, 14, 26] is applied to classify the similarity vectors from each graph (segment) into one of the six sleep stages.

E. Performance Assessment

In this study, the cross validation, sensitivity, kappa coefficient and confusion matrix are used to evaluate the performances of the proposed algorithm.

- *k*-cross-validation: in pattern recognition, *k*-cross-validation is a very popular measure to evaluate the performance of a classification method. It is used to estimate the quality of a classification method by dividing the number of the correctly classified results by the total of the cases. A dataset is divided into *k*-alternately exclusive subsets of an equal size. One subset is used as the testing set, while others are considered as the training sets. All the subsets are tested and the accuracy of the classification is calculated. In this work, the 10-cross-validation is used. The average of the overall results for the subset testing is computed.

$$\text{Performance}=\frac{1}{10}\sum_{1}^{10} \text{accuracy}^{(k)} \qquad (6)$$

where accuracy$^{(k)}$ is the accuracy for the *k*th iteration (*k*=1, 2,….10).

- Sensitivity: is a statistical measure which is used to evaluate the performance of a classification test by measuring the proportion of the actual positive classification. It is defined as

$$\text{Sensitivity} = \frac{TP}{TP+FN} \qquad (7)$$

where TP= the number of the positive classification for all the subset testings decisions, FN= the number of the incorrectly classified decisions for the subsets.

- Confusion matrix: is to provide the information about the actual and predication classification results done by the algorithm.

- Classification accuracy: is the number of the correctly classified decisions divided by the total number of the cases.

- Kappa coefficient: is the measurement of performance agreement of two classification products (the proposed method and the experts in this paper) [33, 43].

V. EXPERIMENTAL RESULTS

To evaluate the proposed method, a set of experiments were designed and conducted using the datasets described in Section III. As mentioned before, different sets of different number of statistical features were used and then were transferred into a graph. A number of experiments were also conducted to determine the best number of the features for each sleep stage based on structural graph similarity properties. The results are discussed in the next section. According to the experimental results, the proposed method achieves a 95.4% accuracy and outperforms the other four existing methods. The experiments were conducted using Matlab (Version: R2013) on a computer with 3.40 GHz Intel(R) core(TM) i7 CPU processor machine, and 8.00 GB RAM.

A. Classification accuracy with different number of features

From the experiment results, we notice that there is a positive relation between the number of the extracted features and the classification accuracy. When the graph nodes increase due to the increase of the number of the extracted statistical features, the discriminative characteristics among the graphs improve, showing more significant differences among sleep stages.

In multi sleep stages classification such as the six sleep stages, it is difficult to obtain high accuracies with the same set of features [42, 43, 13, 19]. In this study, the SGSKM method was also conducted with different sets of statistical features to extract the key characteristic data from the graphs to yield high classification accuracies for six sleep stages.

The SGSKM method was conducted with two datasets to investigate the performances with different channels and a different data size. To determine which set of features can best represent the original EEG signal the best, several experiments were conducted with different sets of features (different number of graph nodes). The 12 statistical features of $\{X_{Me}, X_{Max}, X_{Min}, X_{Mean}, X_{Mod}, X_{Rand}, X_{Q1}, X_{Q2}, X_{SD}, X_{Var}, X_{Ske}, X_{Kue}\}$ were extracted. They were then tested and conducted, separately, to evaluate its classification accuracy from the proposed algorithm. After the proposed method was repeated for 12 times, the importance of each feature was decided based on the classification accuracy results. All the features were then sorted in descending

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2016.2552539, IEEE Transactions on Neural Systems and Rehabilitation Engineering

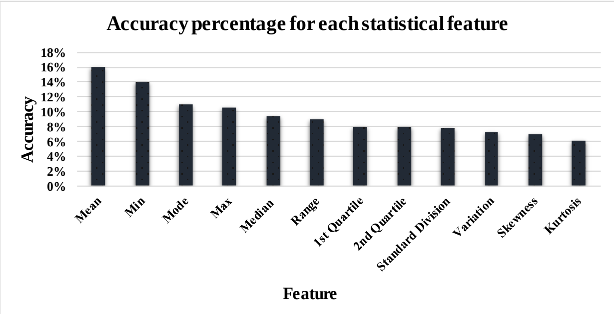


Fig. 4. Order of the statistical features based on the classification accuracy

order based on their classification accuracy importance. The final obtained vector based on the results is $\{X_{Mean}, X_{Min}, X_{Mod}, X_{Max}, X_{Me}, X_{Rand}, X_{Q1}, X_{Q2}, X_{SD}, X_{Var}, X_{Ske}, X_{Ku}\}$. Fig. 4 shows that $\{mean\}$ feature is the most important one among the 12 features, with an average accuracy of 16.5% for dataset-1, and 15.7% for dataset-2.

Six-feature set: the first six features of $\{X_{Mean}, X_{Min}, X_{Mod}, X_{Max}, X_{Me}, X_{Rang}\}$ were used to evaluate the performance of the proposed method. Each segment was represented by a vector of the features containing 450 data points (75X6). The structural graph properties stated in Section IV.B were calculated from each graph. It can be noticed that, the performance of the proposed method significantly improved for several specific sleep stages after having increased the graph nodes with more features. The local graph connectivity significantly increased for stages of: AWA, S1, S2, based on the degree distribution and the average of clustering coefficients. However, the six features were not enough to distinguish the stages: S3, S4 and REM, with a high accuracy, due to the overlapping in similarities of the graph nodes for these stages. Table 4 provides a summary of the obtained results for the two datasets based on the six features set.
The average of 10-cross validation were 7% and 53% with the six features set for dataset-1 and dataset-2 respectively. In order to further increase the accuracy, we increased the number of the features to ten in the next experiment.

Ten-feature set: Top ten features $\{X_{Mean}, X_{Min}, X_{Mod}, X_{Max}, X_{Me}, X_{Rang}, X_{Q1}, X_{Q2}, X_{SD}, X_{Var}\}$ were selected to represent the EEG data. From each segment, 10x75 statistical features from all sub-segments, were extracted and then were mapped into a graph. A very interesting observation in this experiment is that the accuracies by the proposed method for all the six sleep stages significantly improved when the features were increased to ten. The classification accuracies for stage: AWA, S1 and S2 exceeded 85%, increased by more than 25% from the previous experiment with the six feature set for the both datasets. Also, the performances of the proposed method were elevated to 69% for the stages of S3, S4 and REM. Based on the results, the differences in the accuracies between the stages group of AWA, S1 and S2, and the group of the stages of S3, S4 and REM, indicated that it was difficult to use the same set of features to classify all six sleep stages.

TABLE 4
CLASSIFICATION ACCURACY BASED ON SIX FEATURES

Sleep Stages	Dataset-1	Dataset-2
AWA	60.5%	56.5%
S1	59%	55.7%
S2	58.8%	57.5%
S3	53%	52.4%
S4	58%	55.7%
REM	53.7%	50.8%
Average	57.2%	54.8%

TABLE 5
CLASSIFICATION ACCURACY BASED ON TEN FEATURES

	Dataset-1		Dataset-2	
Sleep Stages	Accuracy Rate	Kappa coefficient	Accuracy Rate	Kappa coefficient
AWA	86%	0.36	85%	0.31
S1	88%	0.46	89.6%	0.46
S2	87%	0.25	86.8%	0.29
S3	69%	0.20	68%	0.19
S4	69%	0.21	71.2%	0.20
REM	70%	0.23	73.8%	0.25
Average	78.1%	0.29	79.1%	0.28

Table 5 presents the classification accuracies for the two datasets. It can be seen that the proposed method yields a high performance by using the ten features. There are no significant differences in the results obtained for both datasets. Also, the kappa coefficients for stages: AWA, S1, S2, are very close for the two datasets.
The averages of classification accuracies of the proposed method for 10 times for dataset-1 and dataset-2 are 79% and 75% respectively. From the experimental results for the both datasets, it is clear that the 10 features set is sufficient to classify the stage of: AWA, S1 and S2 from the six sleep stages. The results in Table 5 demonstrates that the ten features set yields an improved performance with the stage of: AWA, S1 and S2, for the different channel datasets, compared with stages: S3, S4, and REM.

Twelve features: In this case, all the twelve features of $\{X_{Mean}, X_{Min}, X_{Mod}, X_{Max}, X_{Me}, X_{Rang}, X_{Q1}, X_{Q2}, X_{SD}, X_{Var}, X_{Ske}, X_{Ku}\}$ were used to classify the six sleep stages. From each segment 900 features were extracted and were transferred to a graph. Table 6 shows the classification results and kappa coefficients for the six sleep stages with the two datasets.
The most noticeable results in this experiment were that the accuracies for all the six sleep stages were exceeded 90% for the both datasets. According to the obtained results in Table 6, the accuracies for the stages of: S3, S4 and REM, improved by more than 19% compared to the previous results in Table 5. More importantly, the results proved that the features {skewness, kurtosis} have significantly increased the accuracy of the classification results for the

TABLE 6
CLASSIFICATION ACCURACY BASED ON TWELVE FEATURES

	Dataset-1		Dataset-2	
Sleep stages	Accuracy rate	Kappa coefficient	Accuracy rate	Kappa coefficient
AWA	97%	0.96	95%	0.92
S1	96 %	0.95	94.6%	0.77
S2	97%	0.92	96.5%	0.95
S3	94.34%	0.67	94%	0.63
S4	95.44%	0.48	95%	0.47
REM	95.77%	0.96	95.8%	0.97
Average	95.93%	0.82	95.15%	0.79

stages (S3, S4 and REM). Based on the results, the stages: AWA, S2, S1 and REM, recorded the highest accuracy compared with the other stages in the two datasets. Also, the kappa coefficient of the six sleep stages increased nearly more than half in the two datasets compared with the previous results in Table 4. The averages of 10 times classification accuracies for dataset-1 and dataset-2 were 95% and 94% respectively Tables 7 and 8 illustrate the confusion matrix and sensitivity for the six sleep stages classification. From the obtained results there are no significant differences in the classification accuracy between the two datasets although the size and the channel of the two datasets are different. The sensitivity of AWA and REM in Tables 7 and 8 is 99.1%, 98.7% and 89.3%, 87.1% respectively, which reveals that the proposed method for Pz-Oz and C3-A1 channels is effective to classify AWA and REM stages from the six sleep stages using the twelve features. Fig.5 shows the classification accuracy against the number of the features used in the classification. It can be noticed that the twelve features set yields the highest accuracy for the six sleep stages with both datasets.

B. Comparative Study

To investigate the performances of the proposed method, we made two types of performance comparisons in this section. Firstly, the performance of the proposed method was compared with the support vector machine. Secondly, the obtained results in Table 5 were compared with other four existing methods from the literature. For a fair comparison, we evaluated the proposed method with other studies that used the same datasets and the same EEG channels. The obtained results for dataset- 1 were listed in Table 6 and were used in the comparison study. There were no reported experimental results for dataset-2.

a. Comparison with the SVM

In this section, the comparison was made in terms of the classification accuracy. The 12 statistical features extracted in Section IV.A were used for the comparison. For the classification accuracy, Fig. 6 presents the comparison between the SVM classifier and the proposed method. The same number of segments were used in the both methods. The segments were selected randomly from dataset-1 and dataset-2.

TABLE 7
CONFUSION MATRIX AND SENSITIVITY OF SIX SLEEP STAGES (DATASET-1)

		Expert's Scoring					
		AWA	S1	S2	S3	S4	REM
The proposed method	AWA	780	10	120	4	9	24
	S1	10	484	36	15	47	8
	S2	20	18	3612	4	0	50
	S3	15	78	89	578	3	75
	S4	9	0	90	52	547	14
	REM	12	14	95	19	21	1458
Sensitivity		99%	80%	88%	86%	87%	89%

TABLE 8
CONFUSION MATRIX AND SENSITIVITY OF SIX SLEEP STAGES (DATASET-2)

		Expert's Scoring					
		AWA	S1	S2	S3	S4	REM
The proposed method	AWA	8750	120	137	2	45	12
	S1	18	2765	78	63	17	7
	S2	78	10	19030	3	611	5
	S3	2	9	129	3216	247	95
	S4	3	17	850	245	3504	921
	REM	7	85	960	114	17	8000
Sensitivity		98.7%	83%	89.8%	86%	79%	87%

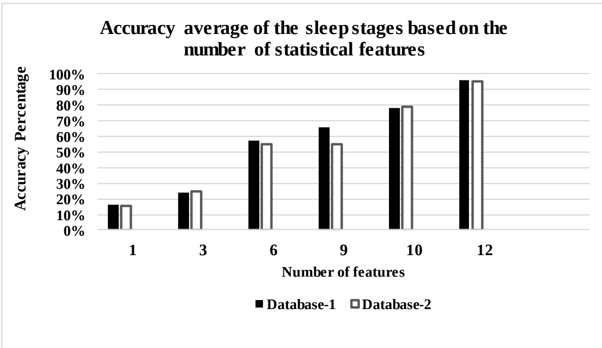


Fig.5. Classification accuracy based on the number of statistical features

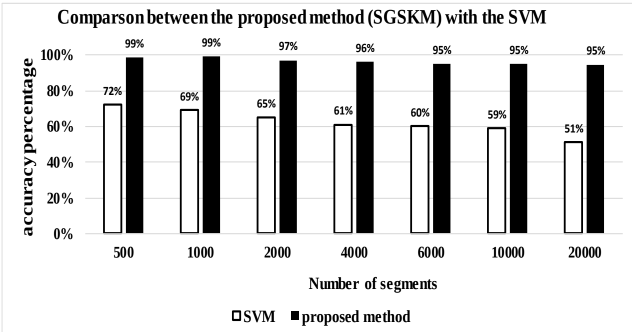


Fig. 6. Accuracy comparison between the proposed method and the SVM

The selected segments were divided into two sets, the training and testing sets. Then, the extracted features (12) from each segment were forwarded to the SGSKM and the SVM. From Fig. 6, it can be noticed that the performance of the proposed method is better than that of the SVM. The accuracy of the SVM was degraded from 72% to 51% when dealing with 20000 segments.

TABLE 9
ACCURACY COMPARISON WITH OTHER METHOS

Authors	Sleep stages, No. of segments	Method	Accuracy
Bajaj et al. [2]	AWA, S1, S2, S3, S4, REM (4700)	Smoothed pseudo Wigner Ville distribution based on time frequency representation and segmentation.	92.3%
Zhu et al. [43]	AWA S1, S2, S3, S4 REM (14963)	Difference visibility graph.	87.5%
Aboalayon et al. [1]	AWA, S1 (200)	Designed band pass filter to decomposed EEG signal.	92%
Güneş et al. [11]	AWA, S1, S2, S3, REM (4195)	K-means clustering based features weighting combing with k-nearest neighbour.	82.15%
The Proposed method	AWA, S1, S2, S3,S4, REM (14963)	Statistical features and structural graph similarity.	**95.93%**

In contrast, the accuracy of the proposed method slightly changed and yielded more accurate and stable results, compared with the SVM. The average of 10 times accuracies was calculated for each of the two methods.

b. The comparison with other existing sleep stage classification methods

To evaluate the performance of the proposed method, the classification results were compared with those of other four existing methods. Table 9 presents the performances comparison among the proposed method and four other reported methods (Zhu et al. [43]; Bajaj et al. [2]; Aboalayon et al. [1] and Güneş et al. [11]. All those studies were used the same dataset (dataset-1) as discussed in Section III. The winning classification accuracy rates among the four methods were highlighted in bold font in Table 8. Based on the results in Table 8, the proposed method is the best among the five methods. Bajaj et al. [2] reported their results of the sleep stages classification with the same dataset, but using lesser segments compared with the proposed method. The average of the accuracy results they achieved was 92.3% for the stages of {AWA, S1, S2, S3, S4 and REM}. The research by Bajaj et al. [2] used a smoothed pseudo Winger-Villa distribution to obtain a segmented time-frequency image (TFI). The TFI divided the EEG signal into five sub-images which corresponded to frequency bands. Then, different features from the histogram of the TFI were extracted. The average accuracy rates were lower than that obtained in this study. Zhu et al. [42] focused on designing a system to differentiate sleep stages using difference visibility graphs. The average of the accuracy results obtained was 87.5% with 14,963 segments. The accuracy rates were also lower than that of the proposed method although the number of segments was the same as used in this paper. Aboalayon et al. [1] proposed an efficient method to distinguish between AWA and stage1 by designing infinite impulse response (IIR) Butterworth band filters. Firstly, the Butterworth band filters was utilized to filter an EEG signal. Then, the filtered EEG signal was decomposed into five frequency bands (delta, theta, alpha, beta and gamma). Different statistical characteristics were pulled out from each band and then were fed to a SVM. They achieved an average of 92% accuracy for the stages of, AWA and S1. They used only 200 segments from 8 subjects in the dataset. Based on our results, the proposed method yields the highest accuracy with a huge number of segments compared with Aboalayon et al. [1]. Güneş et al., [11] proposed an automatic sleep stages scoring method using K-means clustering based on weighted features. The Welch method was used to extract 129 frequency domain features from an EEG signal. Then, the features were weighted and reduced to four. The reduced features were forwarded to the K-means clustering algorithm. They used a lesser number of segments compared with that of the proposed method. In addition, they reported a classification accuracy rate less than the accuracy obtained by the human experts, which is 83±3%. We can see from the above results, the proposed method achieved the highest accuracy compared with the four other existing methods using the same different EEG channels, while a huge amount of data were used in this paper.

VI. CONCLUSION

In this paper, the statistical features and structural graph properties are used to classify the sleep EEG signals. We have investigated the classification ability of graphs constructed from different statistical features sets in sleep EEG classification. It is found that the 12 features set yields a better performance for all sleep stages (95.93%), with a 5% improvement compared with the recent studies even the proposed method in this paper is applied to a huge amount of data. It is also observed that using a ten features set gives reasonable classification accuracies for stages AWA, S1, S2. This study suggests that the graph theory combined with time domain features can be used to classify EEG signals efficiently without pre-processing. The proposed method is compared with the four other existing methods and the SVM. It is demonstrated that the proposed method can achieve the best performance in terms of the classification accuracy. This method can help physicians to accurately diagnose and treat sleep disorders. It can also be applied to different medical data types and be used to different application areas. With large amounts of EEG recordings available, we will apply big data technologies to classify EEG sleep stages.

References

[1] K. Aboalayon, H. T. Ocbagabir, and M. Faezipour, "Efficient sleep stage classification based on EEG signals," in Systems, Applications and Technology Conference (LISAT), 2014 IEEE Long Island, 2014, pp. 1-6.

[2] V. Bajaj and R. B. Pachori, "Automatic classification of sleep stages based on the time-frequency image of EEG signals," Computer methods and programs in biomedicine, vol. 112, pp. 320-328, 2013.

[3] V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren, "A measure of similarity between graph vertices: Applications to synonym extraction and web searching," SIAM review, vol. 46, pp. 647-666, 2004.

[4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," Physics reports, vol. 424, pp. 175-308, 2006.
[5] M. R. Daliri, "Kernel earth mover's distance for eeg classification," Clinical EEG and neuroscience, vol. 44, pp. 182-187, 2013.
[6] L. Doroshenkov, V. Konyshev, and S. Selishchev, "Classification of human sleep stages based on EEG processing using hidden Markov models," Biomedical Engineering, vol. 41, pp. 25-28, 2007.
[7] F. Ebrahimi, M. Mikaeili, E. Estrada, and H. Nazeran, "Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients," in Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, 2008, pp. 1151-1154.
[8] Z. Fang and J. Wang, "Efficient identifications of structural similarities for graphs," Journal of Combinatorial Optimization, vol. 27, pp. 209-220, 2014.
[9] L. Fraiwan, K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, and H. Dickhaus, "Classification of sleep stages using multi-wavelet time frequency entropy and LDA," Methods of information in Medicine, vol. 49, p. 230, 2010.
[10] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, et al., "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," Circulation, vol. 101, pp. e215-e220, 2000.
[11] Güneş, K. Polat, and Ş. Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting," Expert Systems with Applications, vol. 37, pp. 7922-7928, 2010.
[12] Y. He and A. Evans, "Graph theoretical modeling of brain connectivity," Current opinion in neurology, vol. 23, pp. 341-350, 2010.
[13] Y.-L. Hsu, Y.-T. Yang, J.-S. Wang, and C.-Y. Hsu, "Automatic sleep stage recurrent neural classifier using energy features of EEG signals," Neurocomputing, vol. 104, pp. 105-114, 2013.
[14] X. Huang and W. Lai, "Clustering graphs for visualization via node similarities," Journal of Visual Languages & Computing, vol. 17, pp. 225-253, 2006.
[15] S. A. Imtiaz and E. Rodriguez-Villegas, "A low computational cost algorithm for rem sleep detection using single channel EEG," Annals of biomedical engineering, vol. 42, pp. 2344-2359, 2014.
[16] K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," ACM computing surveys (CSUR), vol. 31, pp. 264-323, 1999.
[17] Kemp, "The Sleep-EDF Database," World Wide Web, http://www. physionet. org/physiobank/database/sleep-edf/, accessed August, 2009.
[18] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberyé, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG," Biomedical Engineering, IEEE Transactions on, vol. 47, pp. 1185-1194, 2000.
[19] N. Kerkeni, F. Alexandre, M. H. Bedoui, L. Bougrain, and M. Dogui, "Automatic classification of Sleep Stages on a EEG signal by Artificial Neural Networks," in 5th WSEAS International Conference on Signal, Speech and Image Processing-WSEAS SSIP'05, Corfu, Greece 2005.
[20] X. Li, X. Hu, C. Jin, J. Han, T. Liu, L. Guo, et al., "A Comparative Study of Theoretical Graph Models for Characterizing Structural Networks of Human Brain," International journal of biomedical imaging, vol. 2013, 2013.
[21] A. A. o. S. Medicine and C. Iber, The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications: American Academy of Sleep Medicine, 2007.
[22] S. Micheloyannis, E. Pachou, C. J. Stam, M. Vourkas, S. Erimaki, and V. Tsirka, "Using graph theoretical analysis of multi channel EEG to evaluate the neural efficiency hypothesis," Neuroscience letters, vol. 402, pp. 273-277, 2006.
[23] M. Mourtazaev, B. Kemp, A. Zwinderman, and H. Kamphuisen, "Age and gender affect different characteristics of slow waves in the sleep EEG," SLEEP-NEW YORK-, vol. 18, pp. 557-557, 1995.
[24] M. A. Naderi and H. Mahdavi-Nasab, "Analysis and classification of EEG signals using spectral analysis and recurrent neural networks," in Biomedical Engineering (ICBME), 2010 17th Iranian Conference of, 2010, pp. 1-4.
[25] M. Obayya and F. Abou-Chadi, "Automatic classification of sleep stages using EEG records based on Fuzzy c-means (FCM) algorithm," in Radio Science Conference (NRSC), 2014 31st National, 2014, pp. 265-272.
[26] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," Expert Systems with Applications, vol. 38, pp. 13475-13481, 2011.
[27] F. Panzica, G. Varotto, F. Rotondi, R. Spreafico, and S. Franceschetti, "Identification of the epileptogenic zone from stereo-EEG signals: a connectivity-graph theory approach," Frontiers in neurology, vol. 4, 2013.
[28] Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," Washington, D.C: Government Printing Office, 1968.
[29] J. L. Rodríguez-Sotelo, A. Osorio-Forero, A. Jiménez-Rodríguez, D. Cuesta-Frau, E. Cirugeda-Roldán, and D. Peluffo, "Automatic Sleep Stages Classification Using EEG Entropy Features and Unsupervised Pattern Analysis Techniques," Entropy, vol. 16, pp. 6573-6589, 2014.
[30] M. Ronzhina, O. Janoušek, J. Kolářová, M. Nováková, P. Honzík, and I. Provazník, "Sleep scoring using artificial neural networks," Sleep medicine reviews, vol. 16, pp. 251-263, 2012.
[31] J. T. Sarsoh and K. M. Hashem, "Classifying of Human Face Images Based on the Graph Theory Concepts," Global Journal of Computer Science and Technology, vol. 12, 2012.
[32] S. E. Schaeffer, "Graph clustering," Computer Science Review, vol. 1, pp. 27-64, 2007.
[33] Suily, Y. Li and P. P. Wen, "Clustering technique-based least square support vector machine for EEG signal classification," Computer methods and programs in biomedicine, vol. 104, pp. 358-372, 2011.
[34] Şen, M. Peker, A. Çavuşoğlu, and F. V. Çelebi, "A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms," Journal of medical systems, vol. 38, pp. 1-21, 2014.
[35] J. Stam and J. C. Reijneveld, "Graph theoretical analysis of complex networks in the brain," Nonlinear biomedical physics, vol. 1, p. 3, 2007.
[36] S. Sun and C. Zhang, "Adaptive feature extraction for EEG signal classification," Medical and Biological Engineering and Computing, vol. 44, pp. 931-935, 2006.
[37] University of MONS-TCTS laboratory:The dream subjects database(2013), http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseSubjects/
[38] C. van Straaten and C. J. Stam, "Structure out of chaos: Functional brain network analysis with EEG, MEG, and functional MRI," European Neuropsychopharmacology, vol. 23, pp. 7-18, 2013.

[39] Vural and M. Yildiz, "Determination of sleep stage separation ability of features extracted from EEG signals using principle component analysis," Journal of medical systems, vol. 34, pp. 83-89, 2010.
[40] X. Zhang and Q. You, "An improved spectral clustering algorithm based on random walk," Frontiers of Computer Science in China, vol. 5, pp. 268-278, 2011.
[41] Zhang and M. Small, "Complex network from pseudoperiodic time series: Topology versus dynamics," Physical Review Letters, vol. 96, p. 238701, 2006.
[42] Zhovna and I. D. Shallom, "Automatic detection and classification of sleep stages by multichannel EEG signal modeling," in Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, 2008, pp. 2665-2668.
[43] G. Zhu, Y. Li, and P. P. Wen, "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal," Biomedical and Health Informatics, IEEE Journal of, vol. 18, pp. 1813-1821, 2014.
[44] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, and F. Chapotot, "Feature selection for sleep/wake stages classification using data driven methods," Biomedical Signal Processing and Control, vol. 2, pp. 171-179, 2007

Mohammed Abdal Hadi Diykh received his BSc Degree in Computer Science from Thi_Qar University, Iraq, in 2003 and MSc Degree in Information Technology from Voronezh State University, Russia, in 2010. He is currently a PhD student in the Faculty of Health, Engineering and Sciences, University of Southern Queensland, Australia. His research interests include biomedical signal analysis, data mining and image processing, graph theory.

Yan Li is currently an Associate Professor of Computer Science in the Faculty of Health, Engineering and Sciences at the University of Southern Queensland, Australia. Her research interests are in the areas of Big Data Technologies, Artificial Intelligence, Biomedical Engineering, and Signal/Image Processing.

Peng (Paul) Wen is an Associate Professor at the School of Mechanical and Electrical Engineering, University of Southern Queensland, Queensland, Australia. He obtained his PhD, M.E. degree and B. E degree from Flinders University of South Australia, Australia, and Huazhong University of Science and Technology, China. His research interests include control, signal processing, modelling, simulation, incremental learning and biomedical instrumentation.