An Efficient Sleep Scoring Method using Visibility Graph and Temporal Features of Single-Channel EEG

Ritika Jain [1] and Ramakrishnan Angarai Ganesan [1]

Abstract— This work proposes a method utilizing the fusion of graph-based and temporal features for sleep stage identification. EEG epochs are transformed into visibility graphs from which mean degrees and degree distributions are obtained. In addition, autoregressive model parameters, Higuchi fractal dimension, multi-scale entropy, and Hjorth's parameters are calculated. All these features extracted from a single EEG channel (Pz-Oz) are fed to an ensemble classifier called random undersampling with boosting technique. Two different approaches i.e. 10-fold crossvalidation and 50%-holdout are utilized to evaluate the performance of the model. Cross-validation accuracies of 91.0% and 97.3%, and kappa coefficients of 0.82 and 0.94 are achieved for 6- and 2-state classifications, respectively, which are higher than those of existing studies.

Clinical relevance— Automatic and reliable sleep stage classification can reduce the burden of sleep experts in analyzing overnight sleep data ($\sim$ 8 hours). It can also assist them to find specific traits of interest such as spindle density, by providing annotated sleep data (hypnogram), thereby eliminating the need for tedious and expensive manual scoring. An accurate 2-state (wake/sleep) classification is also crucial for the patients with disorders of consciousness, where stimulation during wake state is considered more effective than that in sleep state.

I. INTRODUCTION

Sleep is the most crucial aspect of a healthy life and lack of sleep leads to numerous health issues such as depression, anxiety or even death. Thus, the analysis of sleep patterns can help in the early detection of neurological disorders such as dementia, schizophrenia, depression or Parkinson's disease [1]. Currently, the gold standard for the evaluation of sleep is overnight polysomnography (PSG) which records multiple physiological signals such as electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG) and respiration. However, this requires the experts to visually inspect the whole overnight data ($\sim$ 8 hrs) and annotate them based on the scoring guidelines of Rechtschaffen and Kales (R&K) [2] or American Academy of Sleep Medicine (AASM) [3]. R&K defines six sleep stages namely Wake, four non-rapid eye movement (NREM) stages: S1, S2, S3, S4 and a rapid eye movement (REM) stage. Manual scoring is tedious, subjective and expensive. Automated sleep scoring can reduce the time and cost involved. This has resulted in an upsurge in the number of research studies involved in developing a reliable and automatic sleep staging method that can replace the traditional way of sleep scoring. The existing sleep stage classification methods have used a wide variety of features extracted from EEG, EMG or EOG signals

[4, 5, 6, 7, 8]. Most of these studies have used features such as spectral edge frequency, bandpower ratios, wavelet or Wigner-Ville distribution (WVD) to extract both time-frequency information embedded in the signal. Non-linear features such as complexity, fractal dimension, detrended fluctuation analysis, sample entropy, Lyapunov exponent, and teager energy operator have been used by several studies to distinguish between different sleep stages. Further, some studies have also employed neural network and deep learning techniques such as recurrent neural network (RNN), long short-term memory (LSTM) network, convolutional neural network (CNN), transfer learning and attention mechanism approach to classify multiple sleep stages. In recent years, graph theory based time series analysis has drawn a huge attention in neuroscience. Visibility graph (VG), proposed by Lacasa et. al. [9], has been employed by a few studies to analyze ECG or EEG signals. Also, a modified VG called horizontal VG (HVG) is used in some studies for assessing the chaotic characteristics of EEG. Although a lot of work has been done in sleep stage identification, 5-state classification accuracy is still in the range 80%-90% which can definitely be improved [6, 7, 10, 11, 12]. Most works have utilized a large set of features or computationally expensive approaches like deep learning, or multiple signals including EMG, EOG and multi-channel EEG. Hence, there is a scope for improvement by working on any of these issues. In this work, our objective is to provide a single-channel EEG based efficient sleep scoring method by utilizing limited and relevant features that can effectively capture EEG signatures of different sleep stages.

II. MATERIALS AND METHODS

A. Experimental Data

The Expanded Sleep-EDF database is utilized in this study, which is publicly available on Physionet [13]. The recordings consist of horizontal EOG, and two EEG channels i.e. Fpz-Cz and Pz-Oz, sampled at 100 Hz. These are scored on the basis of 30s epochs according to R&K manual. Hence, each epoch comprises 3000 samples. We have considered the EEG recordings of 20 subjects and the corresponding epoch distributions for different sleep stages are shown in Table 1.

TABLE I

Epoch distribution of 20 subjects from Expanded sleep EDF database. S.S.: sleep stage; N.E.: Number of epochs.

S.S.	W	S1	S2	S3	S4	REM	Overall
N.E.	36949	1727	8779	1553	1174	4234	54416

[1]Ritika Jain (email: `ritikajain@iisc.ac.in`) and Ramakrishnan A G (email: `agr@iisc.ac.in`) are with the Department of Electrical Engineering, Indian Institute of Science Bangalore 560012, India
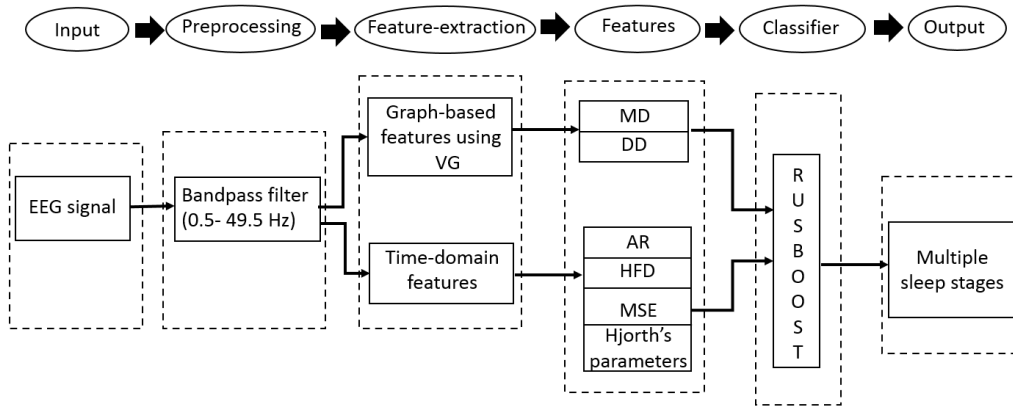
Fig. 1. Flowchart of the proposed method

B. Methodology

The flowchart of the proposed method is shown in Fig. 1. The raw EEG signal is passed through an IIR butterworth bandpass filter with passband frequencies ranging from 0.5 to 49.5 Hz to remove line noise, DC offset and slow drifts. We have segmented the EEG signal into 30s epochs as per the ground truth format. The segments marked with no score or movement time are discarded. Time-domain and graph-based features are extracted from the segmented EEG signals. These features are fed to RUSBoost classifier to perform the classification of multiple sleep stages. The subsections below present a detailed description.

1) Features extracted from visibility graph: Visibility graph (VG) was proposed by Lacasa et. al. [9] to characterize a time series by a graph. Every node in a VG corresponds to a data sample and two nodes are connected if visibility exists between them. In other words, all the samples in between have values less than the value predicted by a linear interpolation. Thus, the visibility criterion is defined as.

$$y_c < y_b + (y_a - y_b)\frac{t_b - t_c}{t_b - t_a} \qquad (1)$$

where (t_a, y_a) and (t_b, y_b) are two arbitrary data points represented as nodes in the graph. These nodes will be visible and hence connected, if another data point (t_c, y_c) placed between them satisfies the above visibility criterion.

For illustration, Fig. 2 shows a time series $x(n) = (0.46, 0.78, 0.39, 0.55, 0.25, 0.65, 0.12, 0.44, 0.29, 0.58)$, where each vertical bar represents a data sample. Here, nodes 2 and 6 are visible and hence connected, since all the data points between them satisfy the visibility criterion. The VG corresponding to the time series is also shown in the figure. The degree of a node is defined as the number of edges connected to it. For instance, the degree of node 2 (representing $x(2) = 0.78$) is 5. We can obtain a degree sequence for this VG by listing the degrees of each node as $(1, 5, 2, 4, 2, 6, 2, 4, 2, 4)$ and mean degree (MD), which is the average of the degree sequence. Another parameter called degree distribution (DD) refers to the probability that a node has degree 'k'. This is obtained as the ratio of the number of nodes with degree 'k' to the total
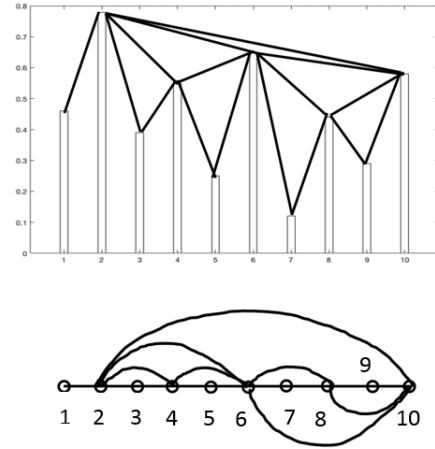

Fig. 2. Conversion of a time series into a visibility graph. The top figure shows an example time series and the connectivity among the individual samples treated as nodes. Here, each vertical bar represents the value of a data sample. Nodes 2 and 6 are visible and hence connected, since all the data points between them satisfy the visibility criterion. The bottom figure shows VG corresponding to the time series.

number of nodes. Here, we denote DD as $P(k)$. In Fig. 2, $P(k) = (0, \frac{1}{10}, \frac{2}{5}, 0, \frac{3}{10}, \frac{1}{10}, \frac{1}{10})$ and MD = 3.2. Since it is less likely for the nodes to have a higher degree, we consider the first few values of $P(k)$ based on its ability to distinguish among different sleep stages. In order to obtain VG from the EEG signal, each epoch is mapped into a graph and its mean degree and degree distributions are computed. The degree distribution $P(k)$ for different values of k is shown in Fig. 3 for one of the subjects. From the figure, it is clear that the degrees ranging from 1 to 5 suffice to distinguish between multiple sleep stages. Hence, we considered MD and $P(k)$ (k from 1 to 5) as the features derived from VG.

2) Features extracted from temporal characteristics: We extracted four time-domain features namely auto-regressive model parameters, Hjorth's parameters, Higuchi fractal dimension and multi-scale entropy from the EEG signal. These features have been selected in particular because they have shown good results in classification problems in many studies, including our previous work [4, 5, 6].

Autoregressive (AR) model: This parametric model represents a sequence as a linear combination of past observations weighted by coefficients $a_1, a_2, ...a_p$ where p is the order of the model. Here, we have used 8th order AR model based on our previous study [4].

$$x[n] = -\sum_{i=1}^{p} a_i x[n-i] + e[n] \qquad (2)$$

Higuchi fractal dimension (HFD): Fractal dimension quantifies the geometrical structures of a signal at multiple scales. Many algorithms exist for estimating it, out of which Higuchi's [14] is the most accurate.

Hjorth's parameters: We have used three Hjorth's parameters i.e. activity, mobility and complexity to assess the statistical properties of the signal in the time domain. All these parameters are computed by using the variance of the signal and its first and second order derivatives [15].

Multi-scale entropy (MSE): This measures the complexity of signals across multiple time scales, which is inherent in biological signals such as EEG and ECG. We have used the sample entropy method for calculating MSE.

$$y_\tau = \frac{1}{\tau}\sum_{i=(k-1)\tau+1}^{k\tau} x_i, \qquad 1 \le k \le \frac{N}{\tau} \qquad (3)$$

The original time series x_i (length N) is converted into a coarse-grained series y_τ by taking the average of all the samples within each non-overlapping window of length τ (refer eqn. 3). Thus, the length of this coarse-grained time series is N/τ whose sample entropy is computed for multiple scales (varying values of τ). Sample entropy involves the parameters r, m and τ, where r is the tolerance of acceptance, m is the dimension of sequence and τ is the scale factor. We use the values of $m = 2$ and $r = 0.15$ times standard deviation, as suggested by Liang [6]. We consider the mean of the MSE values across 3 scales ($\tau = 1, 2$ and 3) since the performance did not improve beyond the value of $\tau = 3$.

3) Classification: In sleep staging, the epochs very uneven distribution across different stages, as seen in Table I. We use the ensemble classifier random undersampling with boosting technique (RUSBoost) for sleep stage classification, since it tackles the class-imbalance problem effectively. The n-class ($n \in \{6, 5, 4, 3, 2\}$) classification problems considered in this study are listed in Table II.

TABLE II
DESCRIPTION OF VARIOUS N-CLASS CLASSIFICATION PROBLEMS USING R&K SCORING.

Classes	Classification criteria
6	W Vs REM vs S1 vs S2 vs S3 vs S4
5	W Vs REM vs S1 vs S2 vs (S3+S4)
4	W vs REM vs (S1+S2) vs (S3+S4)
3	W vs REM vs NREM (S1+S2+S3+S4)
2	W vs Sleep (S1+S2+S3+S4+REM)

III. EXPERIMENTAL RESULTS

We experimented with each of the two available EEG channels (Fpz-Cz and Pz-Oz) and found Pz-Oz to provide a slightly better performance. Hence, all the results reported are for channel Pz-Oz. To evaluate the statistical significance of the features considered, we used Kruskal-Wallis test [16]. All the features have $p < 0.001$ and hence are statistically significant. Fig. 4 shows the box plot of the MDs across multiple sleep stages; MD increases as the subject progresses from lighter to deep sleep, while REM has the least MD value. This is due to the less number of transitions to the REM stage. A similar trend is obtained across all the subjects. Fig. 5 presents the 5-state hypnogram for a subject generated by the proposed method and the experts' score. It can be observed that most of the epochs (2668 out of 2774 epochs=$\sim 96.2\%$) are correctly classified by the proposed method, however there are some spurious jumps at the transitions from one state to another.

Table III shows the confusion matrix between the scores of the expert and our algorithm for 5-state classification with 10-fold cross-validation. It also lists the sensitivity and specificity values for each sleep state. It is evident from the confusion matrix that most of the epochs of N1 stage are misclassified as wake or REM stage, which results in a low sensitivity value ($< 40\%$). The classification accuracy for different classification problems using both 10-fold cross-validation and 50% holdout (averaged over 5 different runs; each run with randomly selected 10 subjects for training and rest 10 subjects for testing) are presented in Table IV. It also compares the performance of our work with some of the existing studies using this dataset. The kappa coefficients (κ) corresponding to n-state classification (n varying from 6 to 2) are 0.82, 0,85, 0.86, 0.89, and 0.94, respectively, indicating an excellent agreement between the scorings provided by our method and those of experts.
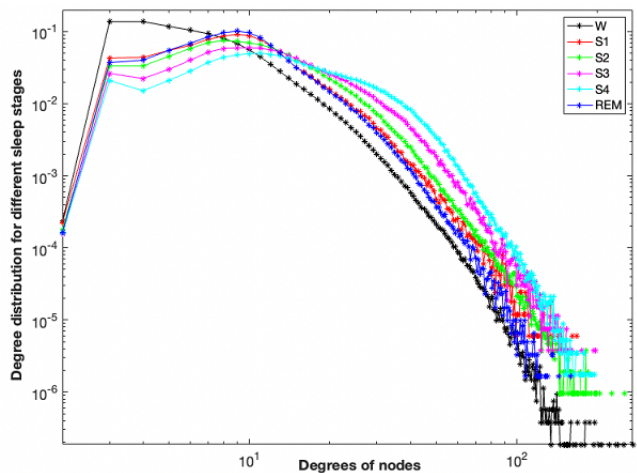

Fig. 3. Log-log plot of degree distribution P(k) across the 6 sleep stages (W, S1, S2, S3, S4 and REM) for a subject as a function of the value of k. black: W; red: S1; green: S2; magenta: S3; cyan: S4; blue: REM.

IV. DISCUSSION

We use the novel time-series analysis method called visibility graph for sleep scoring. The highlight of this work is that a high cross-validation accuracy ($91\% - 97\%$) and κ value ($0.82 - 0.94$) are achieved for different $n-$class (n
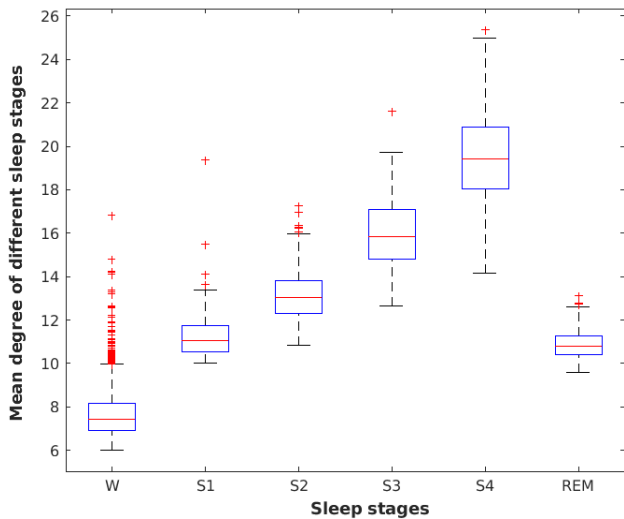
Fig. 4. Box plot showing the 25th, 50th and 75th percentile of MDs over the epochs for the different sleep stages (W, S1, S2, S3, S4 and REM) for a subject. (Outliers shown by '+' and range by the black horizontal lines).

TABLE III

CONFUSION MATRIX, SENSITIVITY AND SPECIFICITY FOR 5-STATE CLASSIFICATION AS PER AASM SCORING (EXP.: EXPERT, ALGO. :ALGORITHM, SENS.: SENSITIVITY, SPEC: SPECIFICITY)

Exp./Algo.	W	N1	N2	N3	REM	*Sens.*	*Spec.*
W	36488	198	76	19	168	98.8	98.1
N1	361	639	252	18	457	37.0	42.7
N2	166	340	7087	571	615	80.7	89.7
N3	14	10	222	2479	2	90.9	80.0
REM	170	308	264	13	3479	82.2	73.7

varies from 6 to 2) classification problems by utilizing a minimal set of features (4 temporal and 2-VG derived) and a single EEG channel. The results indicate the effectiveness of our algorithm for reliable sleep stage identification. The method exploits the knowledge of sleep domain and uses a finite number of features that distinguish between different sleep stages. Hence, it is computationally efficient unlike the deep learning-based methods. A limitation of this study is the low sensitivity for N1 stage which gets misclassified as REM or wake state. Our future work will focus on the improvement of accuracy for N1 sleep stage detection.

REFERENCES

[1] V Latreille et al. "Sleep spindles in Parkinson's disease may predict the development of dementia". In: *Neurobiology of aging* 36.2 (2015), pp. 1083–1090.
[2] A Rechtschaffen. "A manual of standardized terminology, technique and scoring system for sleep stages of human subjects". In: *Public Health Service* (1968).
[3] Conrad Iber et al. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications.* Vol. 1. American academy of sleep medicine Westchester, IL, 2007.
[4] Ramakrishnan Angarai Ganesan and Ritika Jain. "Binary State Prediction of Sleep or Wakefulness Using EEG and EOG Features". In: *2020 IEEE 17th India Council International Conference (INDICON)*. IEEE. 2020, pp. 1–7.
[5] Ramakrishnan Angarai Ganesan and Ritika Jain. "Sleep-awake Classification using EEG Band-power-ratios and Complexity Measures". In: *2020 IEEE 17th India Council International Conference (INDICON)*. IEEE. 2020, pp. 1–6.
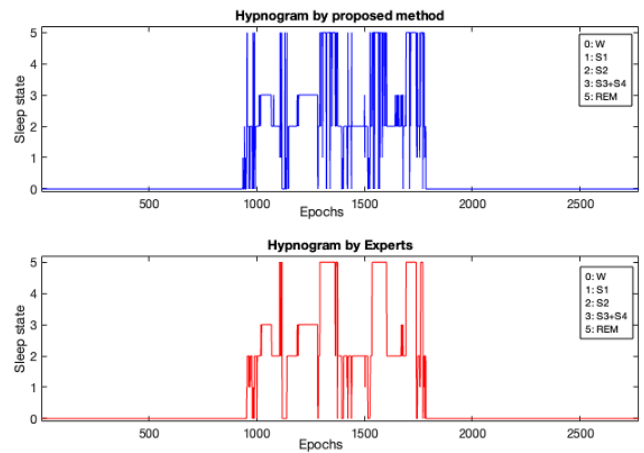
Fig. 5. 5-state hypnogram of a subject generated by the proposed method (blue) and experts' scoring (red)

TABLE IV

COMPARISON OF CLASSIFICATION ACCURACIES (IN %) FOR N-STATE SLEEP IDENTIFICATION PROBLEMS, FOR N=2 TO 6.

Method	Epochs (Subj)	Validation	N=6	N=5	N=4	N=3	N=2
Khalili [10]	41950 (20)	20-fold CV	-	85.4	-	-	-
Liang [6]	15188 (8)	-	-	83.6	-	-	-
Phan [7]	37022 (20)	20-fold CV	-	82.5	-	-	-
Supratak [11]	41950 (20)	20-fold CV	-	82.0	-	-	-
Zhu [12]	14963 (8)	10-fold CV	87.5	88.9	89.3	92.6	**97.9**
Proposed	54416 (20)	10-fold CV	**91.0**	**92.2**	**93.2**	**94.9**	97.3
Proposed	54416 (20)	50% hold-out	86.5	88.4	89.4	90.9	94.6

[6] Sheng-Fu Liang et al. "Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models". In: *IEEE Trans. Instrumentation Measurement* 61.6 (2012), pp. 1649–1657.
[7] Huy Phan et al. "Automatic sleep stage classification using single-channel EEG: Learning sequential features with attention-based recurrent neural networks". In: *40th Annual Inter. Conf. IEEE Eng. in Med. Biol. Society*. 2018, pp. 1452–1455.
[8] Anusha A. S. and Ramakrishnan A. G. "Midline EEG functional connectivity as biomarker for conscious states in sleep and wakefulness". In: *43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2021.
[9] Lucas Lacasa et al. "From time series to complex networks: The visibility graph". In: *Proc. National Acad. Sciences* 105.13 (2008), pp. 4972–4975.
[10] Ebrahim Khalili and Babak Mohammadzadeh Asl. "Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel EEG". In: *Comp. Methods Prog. Biomed.* 204 (2021), p. 106063.
[11] Akara Supratak et al. "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG". In: *IEEE Trans. Neural Sys. Rehab. Eng.* 25.11 (2017), pp. 1998–2008.
[12] Guohun Zhu, Yan Li, and Peng Wen. "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal". In: *IEEE Journal Biomed. Health Informatics* 18.6 (2014), pp. 1813–1821.
[13] A L Goldberger et al. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals". In: *Circulation* 101.23 (2000), e215–e220.
[14] Tomoyuki Higuchi. "Approach to an irregular time series on the basis of the fractal theory". In: *Physica D: Nonlinear Phenomena* 31.2 (1988), pp. 277–283.
[15] B Hjorth. "Time domain descriptors and their relation to a particular model for generation of EEG activity". In: *CEAN-Computerized EEG analysis* (1975), pp. 3–8.
[16] William H Kruskal and W Allen Wallis. "Use of ranks in one-criterion variance analysis". In: *Journal of the Amer. stat. Assoc.* 47.260 (1952), pp. 583–621.