Binary State Prediction of Sleep or Wakefulness Using EEG and EOG Features

Ramakrishnan Angarai Ganesan
Department of Electrical Engineering
Indian Institute of Science
Bengaluru, 560012
Email: agr@iisc.ac.in

Ritika Jain
Department of Electrical Engineering
Indian Institute of Science
Bengaluru, 560012
Email: ritikajain@iisc.ac.in

***Abstract*—Features extracted from the electrooculogram (EOG) or electroencephalogram (EEG) are able to distinguish between waking and sleeping states of a person with 95.2% and 97.5% accuracy, respectively. Sleep EEG data of eight subjects from the Physionet database have been analyzed using 8th order autoregressive model and the model parameters have been used as features, along with Higuchi fractal dimension and sample entropy. Independently, features derived from EOG such as power in a particular band of frequencies and sum of absolute differences were also used. A combination of all these features obtains the maximum accuracy of 98% using 10-fold cross-validation, employing a support vector machine classifier with radial basis function as kernel. Knowing whether a person is awake or asleep is extremely important before presenting sensory stimuli to the patients with disorders of consciousness as a part of neurorehabilitation.**

***Index Terms*—EOG, power, EEG, AR parameters, sleep, wakefulness, coma, SVM, Higuchi fractal dimension, sample entropy.**

I. Introduction

Sleep is a natural mode of relaxation and it has a direct impact on the physical and mental health of human beings. Lack of sleep or sleep deprivation can lead to numerous health problems, depression or even death [2]. Therefore, getting a good quality of sleep is of prime importance. Traditionally, sleep analysis is performed using all-night polysomnographic (PSG) recording which is visually scored by the experts based on Rechtschaffen & Kales's (R&K) scale or American Academy of Sleep Medicine (AASM) rules [10, 14]. Visual scoring is an extremely time consuming and tedious process, and automating it would be extremely useful.

There are basically two types of sleep: rapid eye movement (REM) and non-rapid eye movement (NREM). The latter is further subdivided into three stages, namely N1, N2 and N3. Each of these stages have their associated characteristics. N1 stage is essentially changeover from wakefulness to sleep, with slowing down of heartbeat, breathing, eye movements and even brain waves [2]. As the sleep deepens towards the N2 and N3 stages, all the physiological signals further slow down. N2 stage is characterised by the presence of sleep spindles, K-complexes, or both. N3 is the deepest stage of sleep prominently characterized by delta waves. REM stage is very similar to the awake state, also referred to as the dreaming stage. Utilizing these characteristics, one can distinguish between the different sleep stages. Many researchers are working towards accurate and automatic sleep stage classification [3, 4, 13].

Our aim is to provide a two-stage (sleep or awake) classification rather than multiple stages of sleep such as N1, N2, N3 or REM. We are interested to find whether a patient in coma is asleep or awake. Since there is no specific sleep-wake cycle present in these patients, it is extremely difficult to look for sleep characteristics or patterns in EEG [11]. The best option in this case is to accurately classify these two states (sleep and wake state) of the brain. In this study, we have explored different features derived from EEG and EOG for accurately predicting the sleep/wake states of the brain. We propose a novel algorithm utilizing the non-linear EEG features namely Higuchi fractal dimension and sample entropy, and the autoregressive model along with the spectral features derived from EOG signal to predict the sleep/wake states.

II. Role of EEG and EOG signals in deriving the sleep/wake state

EEG is considered to be the most informative signal for analyzing the different states of the brain such as resting state, sleep or wake state. Information is being processed in all these states but the rate and amount of information varies among the different brain states. Distinct frequency bands of EEG such as alpha (8-13 Hz), delta (0.5-4 Hz) or theta (4-8 Hz) can be used to identify the characteristics of sleep stages. Non-linear dynamics of the EEG signal can be further employed to improve the prediction of sleep/wake states.

EOG signal is frequently used in sleep studies to record the eye movements, which are very prominent in the REM stage of sleep. Spectral features derived from the EOG

978-1-7281-6916-3/20/$31.00 ©2020 IEEE

signal can serve as complementary information that can further improve the prediction accuracy.

III. MATERIALS AND METHOD

Figure 1 presents the flowchart of the proposed algorithm to classify the sleep/wake states.

A. Sleep EEG dataset used for the experiments

In this study, we have used sleep EDF dataset from physionet, which consists of overnight sleep data from 8 subjects, four healthy and four with mild difficulty in falling asleep [5]. These recordings consist of horizontal EOG, Fpz-Cz and Pz-Oz EEG channels sampled at 100 Hz. We have considered EOG and Pz-Oz EEG channel in our study. Each 30 s epoch has been scored by experts as per R & K scale [14].

B. Features computed from EEG and EOG

The raw EEG signal is preprocessed using a bandpass filter with the passband frequencies ranging from 0.1 to 45 Hz. This signal is then segmented into 30s epochs, out of which the unscored epochs are removed. From these segmented epochs, different features are extracted which are provided as input to the classifier for training.

1) Features derived from EEG: Epochs segmented from a single EEG channel Pz-Oz are fed to an AR model which provides AR coefficients as features for each epoch. Further, nonlinear aspects of EEG are captured using Higuchi fractal dimension (HFD) and sample entropy.

Autoregressive (AR) model parameters: An Autoregressive model is used to predict the current behavior of a signal based on its past values. This is one of the most frequently used techniques in linear predictive modelling of time series. The number of samples used for prediction determines the order of the AR model. We have used AR model of order 8 to predict the EEG samples of each individual epoch.

$$x(n) = \sum_{j=1}^{p} a_j x(n-j) \tag{1}$$

where p is the order of the AR model and a_j are AR coefficients, which are used as the derived features.

Higuchi fractal dimension (HFD): Fractal dimension is an important characteristic of a system, because it contains information about the geometrical structures at multiple scales [1]. Higuchi's algorithm [8] is popular for calculating the fractal dimension. In order to obtain the Higuchi's fractal dimension of a non-linear time series, a new set of series is first generated from the original time series, defined as follows:

$$X_k^m : x(m), x(m+k), x(m+2k), ...x(m + \lfloor \frac{N-m}{k} \rfloor k) \tag{2}$$

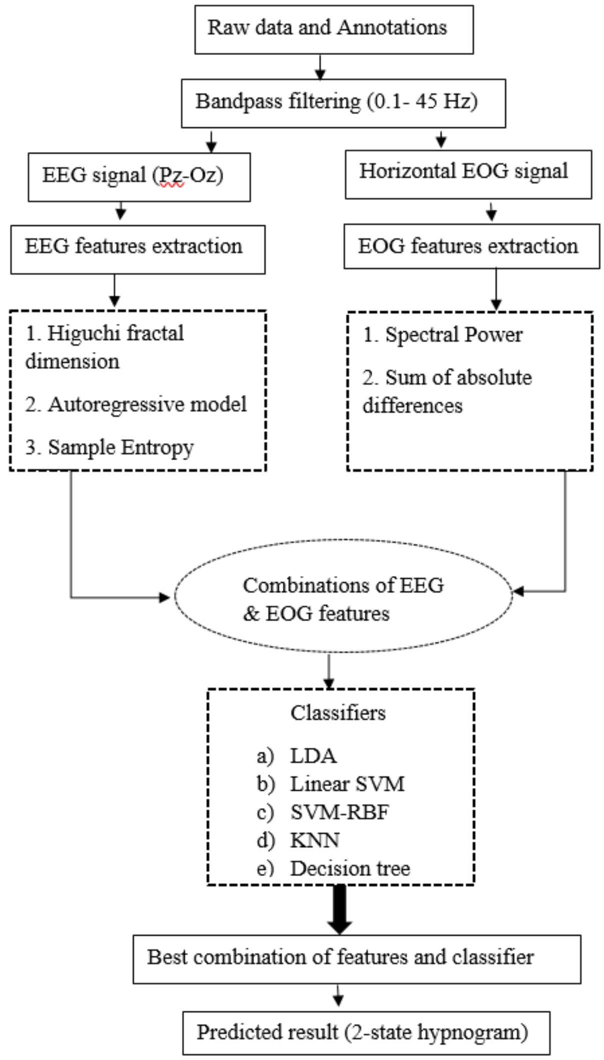

Fig. 1. Flowchart representing the proposed algorithm.

with $m = 1, 2, ...k$. Then the length of the curve associated with each of the newly generated time series X_k^m is calculated as,

$$L_k^m = \frac{1}{k}\Big(\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} \mid x(m+ik) - x(m+(i-1)k) \mid \Big) \Big(\frac{N-1}{\lfloor \frac{N-m}{k} \rfloor k} \Big) \tag{3}$$

$$\langle L(k) \rangle \propto k^{-D} \tag{4}$$

The average value $\langle L(k) \rangle$ of the lengths associated with the time series following a power law given by (4) provides the value of Higuchi fractal dimension 'D'.

Sample entropy (SE): Entropy measures the irregularity of the signal based on the series of patterns embedded in it. Sample entropy, unlike approximate entropy, is easier to implement and also independent of the length of data. Therefore, sample entropy is favored in many studies involving stationary time series analysis [12]. Sample entropy is defined as:

$$S_E(m, r, N) = -\ln\left(\frac{C_{m+1}(r)}{C_m(r)}\right) \tag{5}$$

where C_m is given by

$$C_m(r) = \frac{\{\textit{number of pairs } (i,j) \textit{ with } \mid x_i^m - x_j^m \mid < r, i \neq j\ \}}{\{\textit{number of all probable pairs}\}} \tag{6}$$

The parameters m, r and N are defined as the length of sub-series, tolerance of accepting matches and the total number of samples in the series, respectively. The values of the parameters considered in this study are $m = 2$ and $r = 0.15$ of the standard deviation of the original time series. These values are chosen based on the previous studies [9, 12].

2) Features derived from EOG : EOG signal can effectively measure the eye movements, which form an integral part of REM i.e rapid eye movement sleep stage. Features evaluated from EOG signal can therefore distinguish between the sleep and waking states.

Spectral power (SP): We have calculated the power content of the EOG signal in the frequency band of 0.2-2 Hz. This band is chosen since maximum power is contained within this range. Epochwise power of the signal is obtained from the power spectral density as follows:

$$Power = \sum_{k=k1}^{k2} P_x(k) \tag{7}$$

where P_x is the power spectral density estimated using Welch's method and k1 and k2 are the indices of the DFT corresponding to 0.2 and 2 Hz, respectively.

Sum of absolute differences (SAD): This feature is simple yet powerful to find a trend in the time series. It is defined as the sum of the absolute differences between the consecutive samples in the series.

$$y = \sum_{n=1}^{N} \mid x(n) - x(n-1) \mid \tag{8}$$

where y is calculated for each epoch as the sum of the absolute first difference signal and N is number of samples in an epoch.

C. *Epochwise classification into sleep/waking states*

The EEG and EOG derived features along with the binary state information extracted from the hypnogram are input to different classifier models. We have also compared the performances of different linear and non-linear classifiers: linear discriminant analysis (LDA), support vector machine (SVM) with linear or radial basis function kernel, k-nearest neighbour (KNN) and decision tree. Further, 10-fold cross-validation is applied to evaluate the generalization capability of each model.

IV. Results

A. *Efficacy of EEG derived features in binary state-prediction*

It is evident from Fig. 2 that the first AR coefficient obtained from the EEG accurately captures the sleep/awake state information. Further, Fig. 3 shows that the values of some of the AR coefficients are larger in the waking state than in the sleep state, while others exhibit the opposite trend. This pattern is consistent across all the subjects. The AR coefficients distinctly define the state and provide an accuracy of 96.9% in 2-state classification using SVM-RBF.

Nonlinear features such as sample entropy and Higuchi fractal dimension have been used to capture the non-linearities embedded in the EEG signal. HFD is a measure of self- similarity, and as expected, the fractal analysis enables the identification of these two different states. In wake state, the EEG signal has a fractal behavior with higher values of fractal dimension, while in sleep state, the values get reduced. Similarly, sample entropy, which assesses the signal complexity, is higher for the wake state. It gets reduced as soon as the sleep onset begins. This trend can be seen in all the three EEG features, as shown in Fig 2.

Table I shows the performances in terms of accuracies (in %) of various EEG derived features, both individually, and in different combinations. It also compares the recognition performances of five different classifiers for the different combinations of EEG features. The best feature and classifier combination turns out to be all the three features fed to SVM with RBF kernel. Further, Fig. 4 compares the statistics of the EEG features for their ability to classify sleep/wake states. Thus the order of preference would be the AR coefficients followed by HFD and then sample entropy.

B. *Efficacy of EOG derived features in binary state-prediction*

Figure 5 illustrates that both the SP and SAD features obtained from the EOG characterize the sleep/wake states well. Table II summarizes the classification results for the EOG features using different classifiers. As expected, the best performance is obtained by SVM-RBF classifier for all the features. Individually, they achieve recognition rates of 84.4% and 94.4%, respectively. Combining both the EOG features provides a good accuracy of 95.2% in 2-state classification, without utilizing the information from EEG.

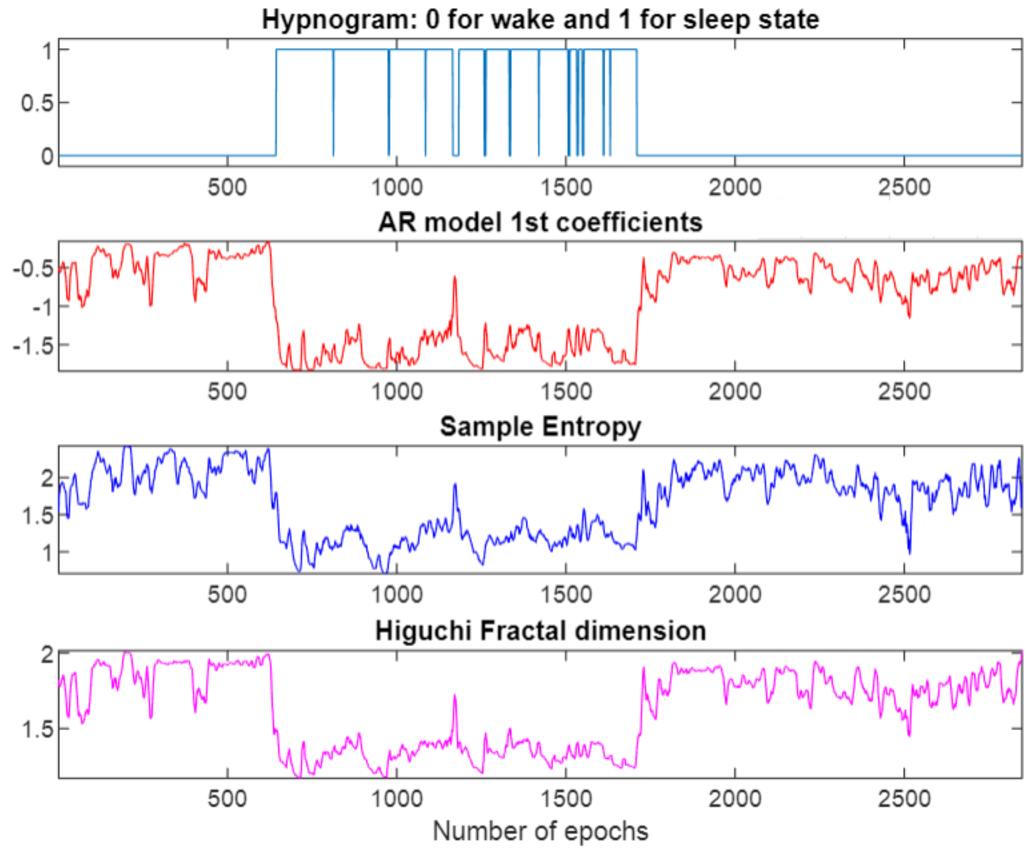

Fig. 2. The hypnogram of subject 1 and the corresponding epochwise values of the EEG features: first AR coefficient, sample entropy and HFD.

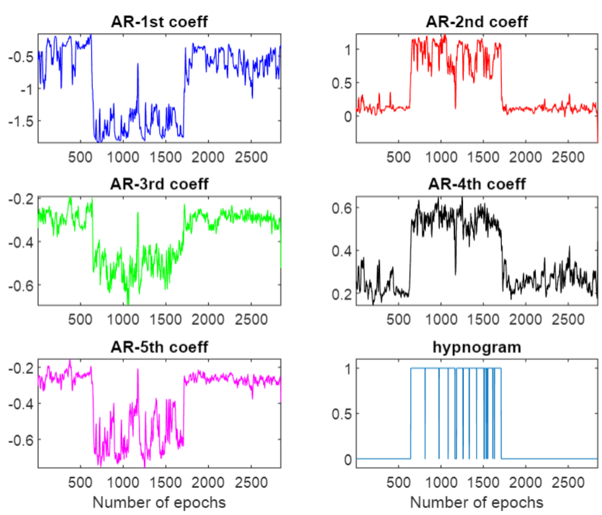

Fig. 3. Plots of the values of the first five AR coefficients and the corresponding hypnogram for subject 1.

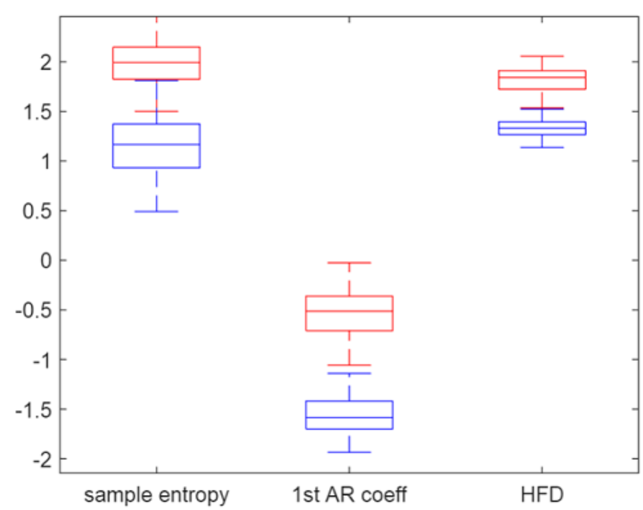

Fig. 4. Boxplot giving median values, 25^{th} and 75^{th} percentiles and range of EEG features during sleep and waking states (red: wake;blue: sleep)

C. Performance of the combined EEG and EOG features

We have already seen that the EOG and EEG features individually provide maximum accuracies of 95.2 and

97.5%, respectively. Table III lists the performances of various combinations of features derived from EEG and EOG using SVM-RBF classifier. The best accuracy of 98.1% is obtained using all the three EEG features and both EOG features.

In order to highlight the potency of the proposed algorithm, we have shown the actual and predicted 2-state hypnograms for two of the subjects. Figures 6 and 7 show that the states of most of the epochs are accurately predicted. Figure 8 shows the confusion matrix between the evaluation of the experts and our algorithm. The sensitivities of wake and sleep states are 98.7% and 97.8%, respectively.

TABLE I
CROSS-VALIDATION RESULTS USING ONLY EEG FEATURES (THE NUMBERS GIVE OVERALL CLASSIFICATION ACCURACIES IN %)

Features	Classification Models				
	LDA	**Linear-SVM**	**SVM-RBF**	**KNN**	**Decision tree**
AR	94.73	95.83	96.89	95.77	94.76
HFD	94.34	94.52	94.56	91.31	91.72
SE	91.45	92.37	92.65	88.29	88.87
AR+HFD	95.74	96.32	97.25	96.26	95.47
HFD+SE	94.36	94.53	94.94	93.07	93.30
SE+AR	95.35	95.89	97.38	96.46	95.46
HFD+SE+AR	95.83	96.33	97.53	96.67	95.75

TABLE II
ASLEEP/AWAKE CROSS-VALIDATION RESULTS (OVERALL EPOCHWISE ACCURACIES IN %) USING ONLY EOG FEATURES

Features	Classification Models				
	LDA	**Linear-SVM**	**SVM-RBF**	**KNN**	**Decision tree**
SP	79.9	81.64	86.2	79.1	80.1
SAD	89.7	93.7	94.4	91.2	91.7
SP+SAD	89.6	94.4	95.2	92.7	93.1

TABLE III
CROSS-VALIDATION RESULTS (EPOCHWISE) USING BOTH EEG AND EOG FEATURES AND THE BEST CLASSIFIER: SVM-RBF

Features	**Accuracy (in %)**
EEG (AR) & EOG (SAD)	97.49
EEG (AR) & EOG (SAD+SP)	97.74
EEG (AR+HFD) & EOG (SAD)	97.85
EEG (AR+HFD) & EOG (SAD+SP)	97.86
EEG (AR+HFD+SE) & EOG (SAD)	97.96
EEG (AR+HFD+SE) & EOG (SAD+SP)	98.08

V. CONCLUSION

Combinations of different features extracted from horizontal electrooculogram and Pz-Oz channel EEG were used with different classifiers to distinguish between waking and sleeping states of the subjects. The best recognition result was obtained with the combination of three features from EEG and two features from EOG using SVM classifier with RBF kernel. The features extracted from EEG are AR parameters, Higuchi fractal dimension and sample entropy. The EOG features are spectral power and sum of absolute differences. The proposed algorithm provides an accuracy of 98.08%, which is higher than the results reported in previous studies on the sleep EDF database for two-state sleep classification [3, 6, 7, 16, 17, 15].

It is significant that only by using the features from a single EEG channel, we can get 97.5% accuracy. Similarly, only by using the features from EOG, one is able to obtain 95.2% accuracy. In the case of patients with disorders of consciousness, one is not sure which signal can be recorded reliably. In such cases, it is useful to be able to have alternate ways (signals) to get good recognition accuracy.

In future, we will work with a much larger database and other features to identify the best feature-classifier combination which can classify the unseen test data with over 99% accuracy.

ACKNOWLEDGMENT

Our sincere thanks are due to Dr. Prathibha Sharan and Mr. Siddharth Vasudevan of NewRo Neuro Rehabilitation Centre, Bangalore for exposing us to the importance of the study of sleep patterns in comatose patients and patients with disorders of consciousness.

REFERENCES

[1] Agostino Accardo et al. "Use of the fractal dimension for the analysis of electroencephalographic time series". In: *Biological cybernetics* 77.5 (1997), pp. 339–350.

[2] Brain Basics. "Understanding sleep". In: *National Institute of Neurol. Disorders and Stroke, Bethesda* (2006).

[3] Christian Berthomier et al. "Automatic analysis of single-channel sleep EEG: validation in healthy individuals". In: *Sleep* 30.11 (2007), pp. 1587–1595.

[4] Luay Fraiwan et al. "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier". In: *Computer Methods and Programs in Biomedicine* 108.1 (2012), pp. 10–19.

[5] 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.

[6] A R Hassan, S K Bashar, and M I H Bhuiyan. "On the classification of sleep states by means of statistical and spectral features from single channel electroencephalogram". In: *International conf. advances in comput. commun. and informatics (ICACCI)*. IEEE. 2015, pp. 2238–2243.

[7] Ahnaf Rashik Hassan and Abdulhamit Subasi. "A decision support system for automated identification of sleep stages from single-channel EEG signals". In: *Knowledge-Based Systems* 128 (2017), pp. 115–124.

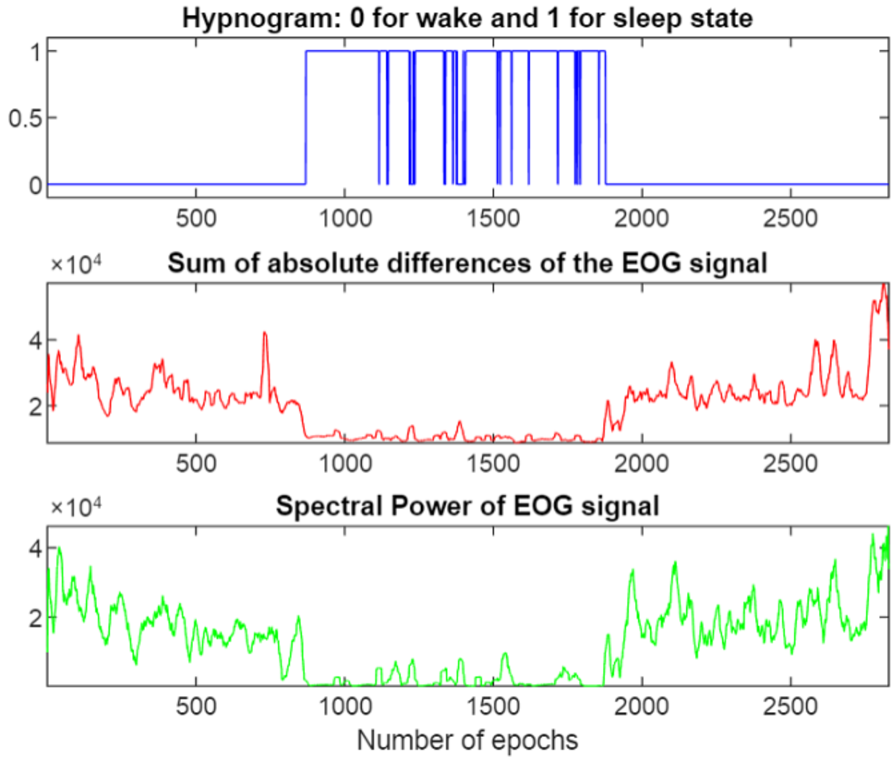

Fig. 5. EOG features: Sum of absolute differences and spectral power of the EOG signal along with the corresponding hypnogram epochwise for subject 2.

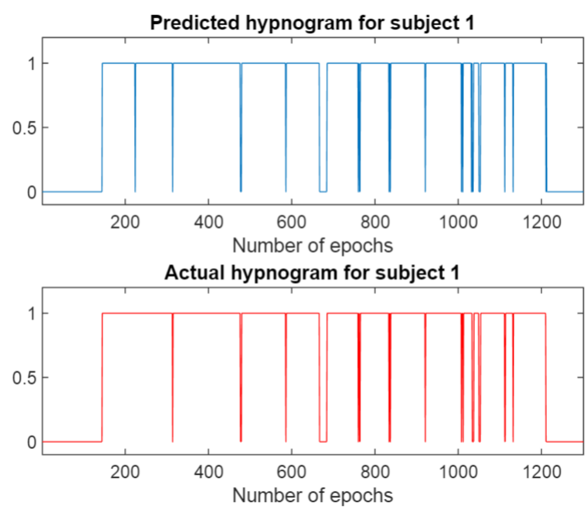

Fig. 6. Hypnogram obtained by the proposed alogrithm and the actual 2-state hypnogram for a healthy subject (0: wake state; 1: sleep state).

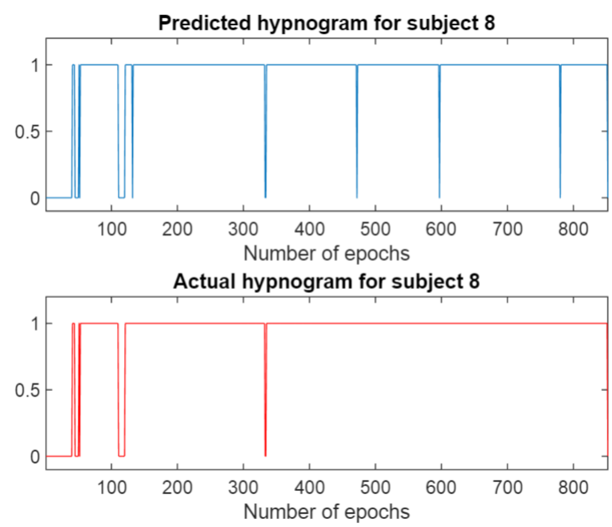

Fig. 7. Hypnogram predicted by our alogrithm and actual 2-state hypnogram for a subject with mild difficulty in falling asleep (0: waking; 1: sleep state).

[8] 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.

[9] Jeng-Rung Huang et al. “Application of multivariate empirical mode decomposition and sample entropy in EEG signals via artificial neural networks for interpreting depth of anesthesia”. In: *Entropy* 15.9 (2013), pp. 3325–3339.

[10] Conrad Iber et al. *The AASM manual for the scoring of sleep and associated events: rules, terminol-*

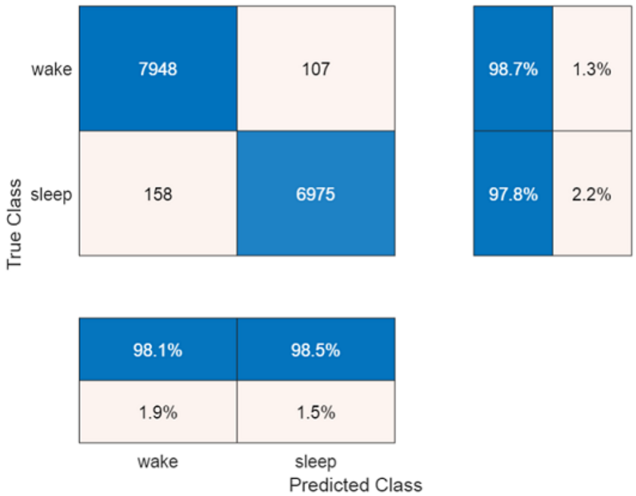

Fig. 8. Confusion matrix between the manual scoring by experts and the scoring by the proposed algorithm for wake and sleep states.

ogy and technical specifications. Vol. 1. American academy of sleep medicine Westchester, IL, 2007.

[11] Ritika Jain and Angarai Ganesan Ramakrishnan. "Electrophysiological and Neuroimaging Studies–During Resting State and Sensory Stimulation in Disorders of Consciousness: A Review". In: *Frontiers in Neuroscience* 14 (2020), p. 987.

[12] George JA Jiang et al. "Sample entropy analysis of EEG signals via artificial neural networks to model patients' consciousness level based on anesthesiologists experience". In: *BioMed research international* (2015).

[13] Sathvik S Prabhu, Neelam Sinha, et al. "Sleep EEG analysis utilizing inter-channel covariance matrices". In: *Biocybernetics and Biomedical Engineering* 40.1 (2020), pp. 527–545.

[14] Allan Rechtschaffen. "A manual of standardized terminology, technique and scoring system for sleep stages of human subjects". In: *Public Health Service* (1968).

[15] Jain Ritika and Angarai Ganesan Ramakrishnan. "Sleep-awake Classification using EEG Band-power-ratios and Complexity Measures". In: *IEEE INDICON*. 2020.

[16] Marina Ronzhina et al. "Sleep scoring using artificial neural networks". In: *Sleep medicine reviews* 16.3 (2012), pp. 251–263.

[17] G Zhu, Y Li, and P Wen. "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal". In: *IEEE journal of biomed. health informatics* 18.6 (2014), pp. 1813–1821.