Single EOG channel performs well in distinguishing sleep from wake state for both healthy individuals and patients

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

***Abstract*— Using a single EOG channel, sleep-wake states of patients with different sleep disorders are accurately classified. We used polysomnography data of 27 patients (mixed apnea, periodic limb movement syndrome, sleep apnea-hypopnea syndrome, and dyssomnia) from DRMS-PAT and 20 healthy subjects from DRMS-SUB databases. We extracted a 67-dimensional feature vector, involving statistical features derived from ensemble empirical mode decomposition, approximate entropy, and relative powers in different frequency bands. Of these, the most relevant features are selected by exploiting mutual information between the features and corresponding labels. RUSBoost classifier is deployed to take care of the unbalanced data distribution. We achieved a high sensitivity of 97.5% and 95.3% as well as high specificity of 96.4% and 93.3% for sleep state in healthy and patients' groups, respectively. Ten-fold crossvalidation accuracies of 91.6% and 95% are achieved for patients and healthy individuals, respectively, using a single EOG channel.**

***Clinical relevance*— Accurate detection of sleep-wake states is crucial for the diagnosis of various sleep disorders including apnea-hypopnea syndrome and insomnia. Automated sleep-wake classification using EOG facilitates easy and convenient long-term sleep monitoring of patients without disturbing their sleep, thereby assisting the clinicians to analyze their sleeping patterns.**

I. INTRODUCTION

Sleep plays a vital role in the development of brain, synaptic pruning, plasticity, memory consolidation and learning [1]. Sleep deficiency and poor quality of sleep cause adverse effects on the mental and physical health of the individual, leading to numerous health problems. Sleep disorders such as insomnia, sleep apnea, and narcolepsy have become prevalent [2] and diagnosis relies on the manual staging of overnight polysomnography (PSG) signals. Manual scoring is time-consuming, tedious and expensive. Further, the use of multiple physiological channels in PSG recordings results in discomfort and hampers the natural sleep of the subject.

Automated sleep stage classification can provide an objective, efficient and faster way of scoring, thereby reducing the burden of experts. Hence, a lot of research is driven towards automatic sleep stage classification using a single channel. Approaches such as time-frequency analysis, non-linear analysis, autoregressive modeling, graph-theory, and spectral analysis have been employed for the classification of multiple stages of sleep with varying degrees of success [3–8]. The performance of any method depends largely on the choice of features. Further, utilizing an appropriate feature selection technique can improve the classification accuracy, while reducing the dimension of the feature space.

Generally, sleep is divided into multiple stages: REM (rapid eye movement), NREM (N1, N2, N3) and wake stage as per the standard scoring criteria of Rechtschaffen & Kales (R&K) [9] or American Academy of Sleep Medicine (AASM) [10]. However, this work focuses only on two-stage classification i.e. sleep (all NREM and REM stages merged) and wake stages. Accurate identification of sleep and wake states are extremely crucial for the diagnosis as well as prognosis of comatose patients [11]. Also, it would help in the timely diagnosis of various sleep disorders like dyssomnia or apnea. The highlights of the study are: (1) mutual information based feature selection to derive optimal set of features for distinguishing between sleep and wake states; (2) utilizing only a single EOG channel; (3) analysis of the performance of the method on patients with various sleep disorders; and (4) adaptive data-driven approach by utilizing ensemble empirical mode decomposition (EEMD).

II. MATERIALS AND METHODS

A. Dataset

We have utilized two publicly available datasets, namely DREAMS Subjects (DRMS-SUB) and DREAMS Patients (DRMS-PAT) dataset [12]. The DRMS-SUB dataset consists of whole-night PSG recordings from 20 healthy volunteers (16 female and 4 male) with mean age of 33.5 ± 14.6 years. The PSG recordings include three EEG channels (CZ-A1 or C3-A1, FP1-A1 and O1-A1), one submental (chin) EMG channel and two EOG channels (EOG1 and EOG2) with 200 Hz sampling rate. The second dataset includes PSG recordings from 27 patients (10 female and 17 male) with mean age of 50.8 ± 14 years having different pathologies such as mixed apnea, periodic limb movements of sleep (PLMS), sleep apnea-hypopnea syndrome (SAHS) or dyssomnia. This dataset has recordings from the same channels as in DRMS-SUB dataset with 200 Hz sampling frequency. Both these datasets were acquired in a sleep laboratory using a digital 32-channel polygraph (Brainnet TM System of MEDATEC, Belgium).

The distribution of epochs of sleep and wake stages for both the datasets is listed in Table I. Also, the number of epochs in wake and sleep state across different patient groups are shown in Table II. In this work, we have utilized only one EOG channel (EOG1) for sleep-wake classification of both healthy and patient groups.

*This work was not supported by any organization

[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

B. Methodology

We utilized EEMD, which is a variant of empirical mode decomposition (EMD) introduced mainly to deal with the mode mixing problem in EMD. EEMD proposed by Wu et. al. [13] involves perturbation of the input by adding random noise and implementing EMD on the noisy input signal. The final intrinsic mode functions (IMFs) are obtained by averaging across the modes of all realizations or ensembles. This approach provides a considerable improvement over the original EMD and is a noise-assisted adaptive data-driven method to decompose the signal into finite mode functions. We used 0.2 as the standard deviation of noise for EEMD.

TABLE I

DISTRIBUTION OF THE SLEEP AND WAKE STATE EPOCHS IN THE TWO DATASETS USED FOR THE STUDY.

Dataset	Wake	Sleep	Total Epochs
DRMS-SUB	5546	23187	28733
DRMS-PAT	11573	30083	41656

TABLE II

NUMBER OF SLEEP AND WAKE STATE EPOCHS ACROSS DIFFERENT PATIENT GROUPS IN DRMS-PAT DATASET

Patient group	Wake	Sleep	Num of patients
SAHS	4101	9945	9
Mixed Apnea	653	2338	2
PLMS	3362	10280	9
Dyssomnia	2616	6759	6
Not specified	841	761	1

The proposed method is represented in Figure 1. The raw EOG signal is filtered by an 8th order Butterworth filter with a pass band of 0.5-49.5 Hz, followed by segmentation into 20 sec epochs. These epochs of length 4000 samples are decomposed into intrinsic mode functions using EEMD. We have selected the first five mode functions and a residual, totalling six modes for each epoch. Statistical features, namely mean, mode, standard deviation, kurtosis and skewness are calculated for each IMF. Also, each IMF is spectrally analyzed using Hilbert Huang transform. This gives instantaneous frequency and energy of the IMF along with its Hilbert spectrum. From the Hilbert spectrum, we compute relative powers (ratio of individual band power to the total power in 0.5-49.5 Hz) of alpha (8.5-11 Hz), beta (15.5-30 Hz), gamma (30-49.5 Hz), delta (0.5-3.5 Hz), k-complex and spindle bands (0.5-1.5 Hz + 11-15 Hz) and the ratio of alpha to theta band power. From the instantaneous frequency (IF) of each IMF, we calculate median, mean, variance and count of changes in the IF. From the instantaneous energy (IE) of each IMF, we compute the median energy. Further, we have included approximate entropy [14] in order to quantify the irregularity in the signal corresponding to sleep and wake stages. All the features considered in this work are listed in Table III.

The extracted features are sorted based on the mutual information (MI) between the feature and the corresponding label, and the features with larger MI values are given

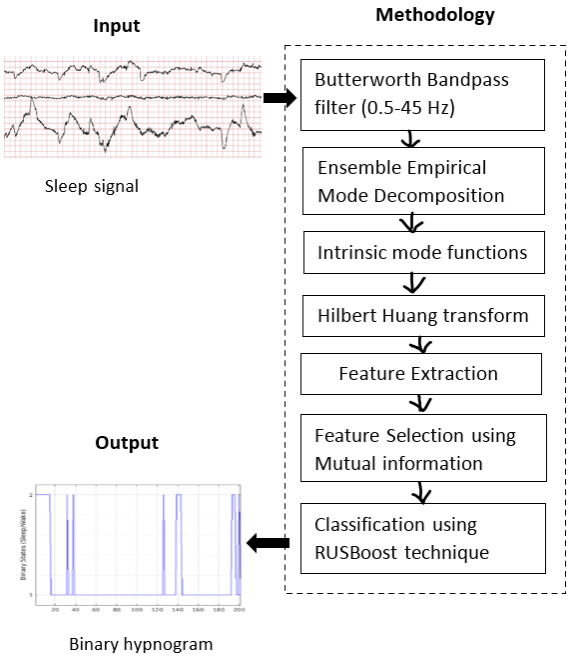

Fig. 1. Flowchart of the proposed method

TABLE III

LIST OF ALL THE FEATURES USED AND THEIR DIMENSIONS.

Features	Dimension	Features	Dimension
Mean	6x1=6	Relative alpha power	1
Mode	6x1=6	Relative beta power	1
Standard deviation	6x1=6	Relative gamma power	1
Kurtosis	6x1=6	Relative delta power	1
Skewness	6x1=6	K-complex + spindle band power	1
Median_IF	6x1=6	Alpha to theta power ratio	1
Mean_IF	6x1=6	Count_IF change	6x1=6
Variance_IF	6x1=6	Appx. Entropy	1
Median_IE	6x1=6	**Total**	67

priority. We have also used the stability test proposed in [15] to choose the optimal set of features. Based on the ranking of features using MI values, top K features (K varied from 15 to 65) are selected in each of the five trials. Each trial randomly selects training and validation samples from the training data. The best subset of features that provide the maximum classification accuracy for the validation data is recorded corresponding to every trial. Then, we count the number of times (i.e. out of 5 trials) a particular feature was present in the best subset. This gives an idea of the stability of the feature; a high value of stability indicates that the feature selection is stable and the particular feature is consistently selected by the model, which confirms its importance. The stability criterion for feature selection is given by,

$$Stability_f = |Count_f(1) - Count_f(0)| \quad (1)$$

where $Count_f(1)$ and $Count_f(0)$ are the number of times (out of five trials) a feature f is present and absent in the subset, respectively. We selected the value of stability as 5 to

decide the optimal feature-set, since it implies that a feature is selected in all the five trials and hence is highly relevant for classification. Further, we used Kruskal-Wallis test (KWT) which is a non-parametric version of one-way ANOVA, to test the statistical significance of features. The features with $p-$value > 0.05 are considered statistically insignificant and removed from the set. The remaining features (after passing the stability test and KWT test) are fed to the classifier to classify an epoch into sleep/wake state.

For classification, we have used random undersampling with boosting technique (RUSBoost) which utilizes an ensemble of decision trees as weak learners. RUSBoost has been shown to outperform traditional SVM mainly due to its ability to deal with class imbalance problem by undersampling the majority class and providing a balanced data to work with [16]. The output of the RUSBoost classifier is a binary hypnogram representing sleep and wake states. In order to evaluate the performance of the proposed method, we have used 10-fold cross validation.

III. Results and Discussion

We have evaluated the suitability of a single EOG channel (EOG1) to classify the sleep/wake states of healthy subjects as well as patients with sleep disorders. The proposed method achieves 10-fold cross-validation accuracy of 95.01% and 91.63% using EOG1 channel on DRMS-SUB and DRMS-PAT datasets, respectively. Table IV summarizes the results of our method for both datasets using 10-fold crossvalidation. High values of 97.5% and 96.4% are achieved for sensitivity and specificity for sleep state detection in healthy subjects. These values are comparatively less ($\sim 84\%$ and $\sim 88\%$) for the wake state. Further, our method also performed quite well on DRMS-PAT dataset. It obtained sensitivity of 95.3% and specificity of 93.3% for the identification of sleep states in the patient population.

We also computed Cohen's kappa values [17] to examine the level of agreement between the experts and proposed method. Kappa values of 0.83 and 0.79 are achieved for DRMS-SUB and DRMS-PAT datasets, respectively, indicating a strong agreement between the scores provided by our method and that of experts. By utilizing the mutual information based feature selection and stability criteria, the dimension of feature-set is reduced from 67 down to 45 for DRMS-SUB dataset. In the case of DRMS-PAT dataset, the dimension of optimal feature-set reduces to 24. This implies that the top 24 features (ranked by MI value) suffice and provide similar classification accuracy as that of all the 67 features. This way we could eliminate the redundant or irrelevant features; which finally result in an increased accuracy with lesser number of features.

Figure 2 shows the test accuracy of five unseen subjects from DRMS-SUB dataset across three different runs. In each run, the model is trained on randomly selected 15 subjects (out of 20), and tested on the remaining 5 subjects. The average prediction accuracies across the five test subjects in three different runs are about 93%, 91%, and 92%, respectively. The prediction accuracies for the unseen test patients from DRMS-PAT dataset are presented in Figure 3. In each run, the model is trained on randomly chosen 22 patients (out of 27) and then tested on the remaining 5 patients. Since the training and testing patients are chosen at random, test data comprises patients with different pathology in each run; which can be seen in Fig 3. On an average, the prediction accuracies across the test patients in the three different runs are about $87\%, 82\%$ and 86%, respectively. These results prove that a single EOG channel is able to provide sufficiently high classification accuracies across both the healthy and patient populations.

TABLE IV

Performance metrics: Accuracy, Kappa value, Sensitivity and Specificity for DRMS-SUB and DRMS-PAT datasets with a single EOG channel (EOG1) using 10-fold crossvalidation

Dataset	Accuracy	Kappa	Sensitivity		Specificity	
			Sleep	Wake	Sleep	Wake
DRMS-SUB	95.01%	0.83	97.5%	84.1%	96.4%	88.4%
DRMS-PAT	91.63%	0.79	95.3%	82.1%	93.3%	87.0%

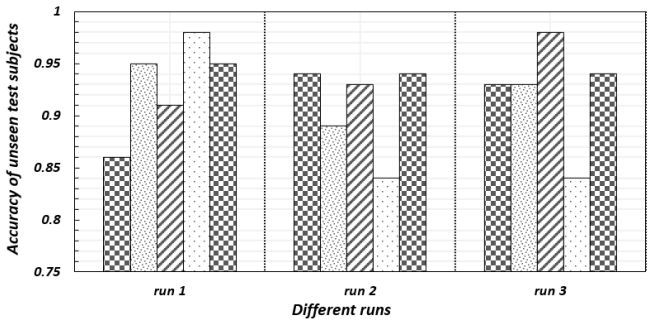

Fig. 2. Accuracy of five unseen healthy test subjects across three different runs for DRMS-SUB dataset using EOG1 channel.

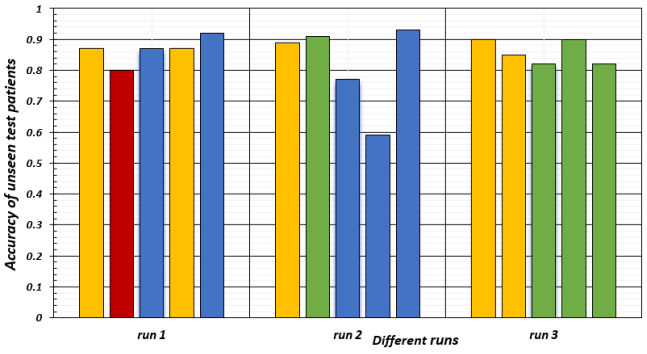

Fig. 3. Accuracy of five unseen test patients with sleep disorders across three different runs (yellow: dyssomnia; red: mixed apnea; green: PLMS; blue: SAHS) for DRMS-PAT dataset using EOG1 channel.

Table V compares the sleep-wake classification accuracy of the proposed method with the existing literature. Since we could not find any study utilizing EOG channel on this dataset, we have compared our method with EEG/EMG/ECG based studies irrespective of the dataset used. Though our

method uses only EOG signal to classify the sleep and wake states, it is able to outperform the results of [18] and achieves performance comparable to other EEG based methods. It provides an average classification accuracy of 95% for $N = 20$ healthy controls and 91.6% for $N = 27$ patients with various sleep disorders.

Very few studies in the literature have explored EOG and among those limited studies, most have used EOG in combination with EEG. It can be seen from the table that though EEG provides the maximum classification accuracy for sleep-wake identification, EOG is able to achieve performance comparable to EEG and far better than other modalities including ECG and EMG. However since the datasets used by various studies are different, it is difficult to directly compare the performance of our method with the others.

TABLE V

COMPARISON OF ACCURACIES (IN %) FOR SLEEP-WAKE CLASSIFICATION ACROSS THE EXISTING LITERATURE

Study	Dataset	Modality	Healthy/Patients	Accuracy
Shen et. al. [3]	DRMS-SUB	EEG	20/Nil	96.2
Ganesan et. al. [4]	Sleep-EDF	EEG+EOG	8/Nil	98.1
Ganesan et. al. [5]	Sleep-EDF	EEG	8/Nil	93.9
MEFF-R [6]	DRMS-SUB	EEG	20/Nil	96.5
Hassan et. al. [18]	DRMS-SUB	EEG	20/Nil	93.3
Brignol et. al. [19]	Private	EEG	7/Nil	94
Adnane et. al. [20]	MIT-BIH	ECG	16/Nil	79.3
Hwang et. al. [21]	Private	EMG (Anterior Tibialis)	7/5	80.5/77.6
Our work	**DRMS-SUB & DRMS-PAT**	**EOG**	**20/27**	**95/91.6**

IV. CONCLUSION

We utilized two recent publicly available datasets i.e. DRMS-SUB for healthy controls, and DRMS-PAT for the patient population to validate the performance of our method. Among the limited studies reporting performance on DRMS-SUB dataset, our method gives comparable classification accuracy with a single EOG channel. To the best of our knowledge, none of the studies have reported classification performance for single channel EOG on patient population. So, this is the maiden work using only an EOG channel with results on patients with different pathologies such as mixed-apnea, PLMS, SAHS, and dyssomnia. This study not only shows the feasibility of using single EOG channel for detection of sleep-wake states in patients, but also confirms its potential to provide results comparable to those of EEG based approaches. This work achieved 2-class classification accuracies of 95% and 91.6% for healthy and patient groups, respectively, using a single EOG channel. An accurate and reliable sleep-wake identification is crucial for the diagnosis of patients in coma and sleep disorders.

The advantage of using single EOG channel for recording overnight sleep data is that it allows an easy and convenient long-term monitoring of sleep. Unlike EEG, it does not require a specialist for electrode placement, and also does not disturb the sleep significantly. An accurate sleep-wake detection method can thus reduce the burden of experts who generally look through the whole data records and search for any irregularity in specific events such as number of arousals or apnea. Our results indicate that the proposed method has potential to assist clinicians in identifying the sleep-wake states in patients. In our future work, we will implement it on a much larger clinical data.

REFERENCES

[1] Charles A Czeisler. "Duration, timing and quality of sleep are each vital for health, performance and safety". In: *Sleep Health: Journal of the National Sleep Foundation* 1.1 (2015), pp. 5–8.

[2] Maurice M Ohayon. "Epidemiology of insomnia: what we know and what we still need to learn". In: *Sleep medicine reviews* 6.2 (2002), pp. 97–111.

[3] Huaming S et al. "An automatic sleep stage classification algorithm using improved model based essence features". In: *Sensors* 20.17 (2020), p. 4677.

[4] Ramakrishnan A G and Ritika Jain. "Binary state prediction of sleep or wakefulness using EEG and EOG features". In: *17th India Council International Conf. (INDICON)*. IEEE. 2020, pp. 1–7.

[5] Ramakrishnan A G and Ritika Jain. "Sleep-awake classification using EEG band-power-ratios and complexity measures". In: *17th India Council International Conf. (INDICON)*. IEEE. 2020, pp. 1–6.

[6] Ritika Jain and Ramakrishnan A G. "Reliable sleep staging of unseen subjects with fusion of multiple EEG features and RUSBoost". In: *Biomedical Signal Processing and Control* 70 (2021), p. 103061.

[7] Ritika Jain and Ramakrishnan Angarai Ganesan. "An efficient sleep scoring method using visibility graph and temporal features of single-channel EEG". In: *43rd Annual International Conf of IEEE Eng Med & Biol Soc (EMBC)*. IEEE. 2021, pp. 6306–6309.

[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 Conf of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2021.

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

[10] Conrad I, Sonia A-I, Andrew L C, et al. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. American Academy of Sleep Med. 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] S Devuyst. *The DREAMS Databases and Assessment Algorithm [Data set]*. 2005.

[13] Zhaohua W and Norden E H. "Ensemble empirical mode decomposition: a noise-assisted data analysis method". In: *Adv in adaptive data analysis* 1.01 (2009), pp. 1–41.

[14] Steven M P. "Approximate entropy as a measure of system complexity." In: *Proceedings of the National Academy of Sciences* 88.6 (1991), pp. 2297–2301.

[15] Fatima K, Mohammed A, and Hashem T. "Efficient feature selection using particle swarm optimization: A hybrid filters-wrapper approach". In: *10th Intern. Conf Info Commun. Sys. (ICICS)*. IEEE. 2019, pp. 122–127.

[16] Chris Seiffert et al. "RUSBoost: A hybrid approach to alleviating class imbalance". In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 40.1 (2009), pp. 185–197.

[17] Mary L McHugh. "Interrater reliability: the kappa statistic". In: *Biochemia medica* 22.3 (2012), pp. 276–282.

[18] Ahnaf R H and Mohammed I H B. "Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting". In: *Computer methods and programs in biomedicine* 140 (2017), pp. 201–210.

[19] Arnaud Brignol, Tarik Al-Ani, and Xavier Drouot. "EEG-based automatic sleep-wake classification in humans using short and standard epoch lengths". In: *12th International Conference on Bioinformatics & Bioengineering (BIBE)*. IEEE. 2012, pp. 276–281.

[20] Mourad Adnane, Zhongwei Jiang, and Zhonghong Yan. "Sleep–wake stages classification and sleep efficiency estimation using single-lead electrocardiogram". In: *Expert Systems with Applications* 39.1 (2012), pp. 1401–1413.

[21] SuHwan Hwang et al. "Sleep/wake estimation using only anterior tibialis electromyography data". In: *Biomedical engineering online* 11.1 (2012), pp. 1–15.