Biomedical Signal Processing and Control 69 (2021) 102898

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Machine learning with ensemble stacking model for automated sleep staging using dual-channel EEG signal

Santosh Kumar Satapathy [a,*], Akash Kumar Bhoi [b,d], D. Loganathan [a], Bidita Khandelwal [c], Paolo Barsocchi [d,*]

[a] Department of Computer Science Engineering, Pondicherry Engineering College, Puducherry 605014, India
[b] Department of Electrical and Electronics Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim 737136, India
[c] Department of General Medicine, Sikkim Manipal Institute of Medical Sciences, Sikkim Manipal University, Gangtok 737102, Sikkim, India
[d] Institute of Information Science and Technologies, National Research Council, 56124 Pisa, Italy

ARTICLE INFO

Keywords:
Sleep stages
Electroencephalogram
Machine learning
Random forest
eXtreme gradient boosting
Ensemble learning stacking model

ABSTRACT

Sleep staging is an important part of diagnosing the different types of sleep-related disorders because any discrepancies in the sleep scoring process may cause serious health problems such as misinterpretations of sleep patterns, medication errors, and improper diagnosis. The best way of analyzing sleep staging is visual interpretations of the polysomnography (PSG) signals recordings from the patients, which is a quite tedious task, requires more domain experts, and time-consuming process. This proposed study aims to develop a new automated sleep staging system using the brain EEG signals. Based on a new automated sleep staging system based on an ensemble learning stacking model that integrates Random Forest (RF) and eXtreme Gradient Boosting (XGBoosting). Additionally, this proposed methodology considers the subjects' age, which helps analyze the S1 sleep stage properly. In this study, both linear (time and frequency) and non-linear features are extracted from the pre-processed signals. The most relevant features are selected using the ReliefF weight algorithm. Finally, the selected features are classified through the proposed two-layer stacking model. The proposed methodology performance is evaluated using the two most popular datasets, such as the Sleep-EDF dataset (S-EDF) and Sleep Expanded-EDF database (SE-EDF) under the Rechtschaffen & Kales (R&K) sleep scoring rules. The performance of the proposed method is also compared with the existing published sleep staging methods. The comparison results signify that the proposed sleep staging system has an excellent improvement in classification accuracy for the six-two sleep states classification. In the S-EDF dataset, the overall accuracy and Cohen's kappa coefficient score obtained by the proposed model is (91.10%, 0.87) and (90.68%, 0.86) with inclusion and exclusion of age feature using the Fpz-Cz channel, respectively. Similarly, the Pz-Oz channel's performance is (90.56%, 0.86) with age feature and (90.11%, 0.86) without age feature. The performed results with the SE-EDF dataset using Fpz-Cz channel is (81.32%, 0.77) and (81.06%, 0.76), using Pz-Oz channel with the inclusion and exclusion of the age feature, respectively. Similarly the model achieved an overall accuracy of 96.67% (CT-6), 96.60% (CT-5), 96.28% (CT-4),96.30% (CT-3) and 97.30% (CT-2) for with 16 selected features using S-EDF database. Similarly the model reported an overall accuracy of 85.85%, 84.98%, 85.51%, 85.37% and 87.40% for CT-6 to CT-2 with 18 selected features using SE-EDF database.

1. Introduction

Maintaining proper health and mental stability is critical for overall health and well-being. Despite a good deal of research investment, sleep quality continues to be a crucial public challenge. Nowadays, people of all age groups are affected by improper sleep quality. Poor sleep can lead to a variety of neurological disorders [1,2]. Sleep disorders are common in all subsets of the population, independently of gender. This public health challenge greatly affects the quality of life in terms of both physical and mental health. Insomnia, parasomnias, sleep-related breathing difficulties, hypersomnia, bruxism, narcolepsy, and circadian rhythm disorders are common examples of sleep-related disorders.

* Corresponding authors.
E-mail addresses: santosh.satapathy@pec.edu (S.K. Satapathy), akash.b@smit.smu.edu.in, akashkrbhoi@gmail.com (A.K. Bhoi), drloganathan@pec.edu (D. Loganathan), bidita.k@smims.smu.edu.in (B. Khandelwal), paolo.barsocchi@isti.cnr.it (P. Barsocchi).

https://doi.org/10.1016/j.bspc.2021.102898
Received 15 February 2021; Received in revised form 20 April 2021; Accepted 19 June 2021
Available online 6 July 2021

Some of these disorders can be treated with proper analysis of early symptoms; in such cases, adequate sleep quality is essential for the patient's recovery. Moreover, numerous sleep disorders can be clinically diagnosed with the help of computer-aided technologies [3]. Sleep monitoring is one of the most significant activities in assessing sleep-related disturbances and other neural problems. Sleep is a dynamic process and includes different sleep stages, including the waking, non-rapid eye movement (NREM), and rapid eye movement (REM) stages. Furthermore, the NREM sleep state is divided into four stages: NREM N1, N2, N3, and N4 [4]. The wake stage is the period of awakening before sleep. The NREM sleep stages are sequentially indicative of light to deep sleep. N1 is a light sleep stage with slow eye and muscle movements. True sleep begins with stage N2, where eye movements stop, and brain activity decreases. The N3 and N4 stages are periods of deep sleep without eye and muscle movements. Finally, in the REM stage, rapid eye movements occur, and breathing increases. A nightly sleep cycle consists of approximately 75% NREM sleep and 25% REM sleep [5]. The asleep test can support sleep assessment with polysomnographic (PSG) recordings. PSG signals are a collection of different physiological signals that are collected from subjects during sleep. A PSG signal is the combination of multivariate signal recordings, electroencephalogram (EEG), electrocardiogram (ECG), electrooculogram (EOG), and electromyogram (EMG) [6]. The EEG signal recordings are used during sleep staging scoring. These signals represent brain activity and, therefore, are suitable for the evaluation of sleep abnormalities. After data collection, an asleep staging score is given. The recorded EEG signals are extracted through multiple fixed electrodes located in different places on a patient's scalp. The process of electrode placement is done according to the international 10/20 placement system [7].

The entire process is carried out by sleep experts who analyze the different patterns of sleep states. The evaluation is made through visual inspection using the recorded data for a specific time window. Consequently, the sleep score is determined through multiple criteria. The criteria for the sleep scoring process are based on the Rechtschaffen Kales guidelines [8]. These guidelines classify the sleep stages as wake (W), non-rapid eye movement (N1, N2, N3, N4), and rapid eye movement (REM). The proposed guidelines also include minor changes introduced by the American Academy of Sleep Medicine (AASM) [9]. The AASM manuals have combined the N3 and N4 stages into a single stage (N3) characterized by slow-wave sleep (SWS). Each sleep stage is directly associated with the different physiological and neuronal behavior, which normally helps analyze the subjects' sleep characteristics. This entire procedure is treated as sleep staging. However, it has been found that the manual sleep staging on the PSG signal is time-consuming, manpower-intensive, and variations on the results concerning the individual experts. Further, the PSG signal is quite inconvenient and uncomfortable for the subjects because its more adhesive electrodes and cabling are fixed in the body during sleep behavior recordings. Henceforth developing a simple and robust automated sleep staging is greatly helpful in this regard. It has been observed that EEG signals are more effective during sleep scoring among all the PSG signals because it directly provides the brain's behavior, which is more helpful for discriminating the sleep behavior with connection to the individual sleep stages.

To design a reliable and convenient automated sleep scoring system, this proposed research work obtains only the single-channel EEG signal to classify the multi-class sleep stages, and it has been reported from the existing contribution that the single-channel EEG signal given promising accuracy results [10,11].

It has been noticed that the major contribution to the sleep scoring methods was based on the traditional shallow learning approaches and EEG signal. It has been found that most of the earlier contributions are followed four basics phases such as data pre-processing, feature extraction, feature selection, and classification. It has been seen that the EEG signal is composed of several signal sub bands such as alpha (α), delta (δ), theta (θ), beta (β), sigma (σ), and k-complex. All these characteristics of waves have appeared in the different sleep stages with different ranges of frequency levels.

A detailed description of the frequency ranges concerning the different characteristics of waves is described in Table 1. From Table 2, it has reported that the σ and δ waveforms are mostly occurred during in N2 and SWS sleep stages, respectively, similarly the α and β wave characteristics are more dominant during the sleep stages in wake and REM and while θ wave pattern is visible during in more N1 sleep stage. Most of the studies [12–14] obtained EEG signal segmentation into the different frequency sub-bands during the pre-processing sleep stages. Lsu et al. [15] considered finite impulse bandpass filter for six sub-bands of the EEG signals with the different ranges of frequency levels such as δ (0.5–2 Hz), sawtooth(2–6 Hz), θ(4–8 Hz), α(8–13 Hz), σ(12–14 Hz), and β(12–30 Hz) to analysis the changes sleep behavior patterns from EEG signals, respectively. Memar P et al. [13] claimed that the significance of the gamma (γ) (30–49.5 Hz) wave has more significance towards the sleep behavior analysis of the subject. It has also been noticed that ignorance of γ wave characteristics may cause degradation of the classification performance. However, all these contributions have ignored the subject's age factor and its effects on the EEG signals. It has been found that there should be a direct link between age and α waveforms. Generating the α waveforms are appeared for three old subjects with a range is of 8 Hz. Similarly, for adults, the same wave patterns are visible with a frequency range of 9–11 Hz. The β wave patterns also dominate with age increases, but it also displays after 60 years. The details of the relationship between age with the frequency range are shown in Table 2. Till date, so many contributions on automatic sleep staging were developed by different researchers with a limited degree of success ratio. It has been seen that the system's performance completely depends on the different features, choice of feature selection algorithms, and finally, the selection of the suitable classifiers. In most studies, the authors have obtained the features like time domain, frequency domain, time–frequency domain, graph features, and empirical mode decomposition features [16,11,17–19].

It has been found that most of the contributions obtain the pre-defined time and frequency domain features without being proper analysis of the feature relevance and redundancy, which directly affects the performance of the classification model and also taken more computation time [19,20]. On the other hand, feature selection algorithms overcome existing drawbacks with the feature extraction by considering the proper analysis of feature relevance and redundancy, which significantly improves the classification performance by selecting the relevant features from the extracted feature vectors [20]. It has found that most of the authors have obtained ReliefF [21], minimum redundancy maximal relevancy (mRMR) [22], Fisher score [23], recursive feature elimination (RFE) [24], Sequential Floating Forward Selection [25], Fast-correlation-Based Filter Solution [26] and information gain [27] used during sleep staging for selecting the suitable features. Sometimes it has also been noticed that the selection of the proper classification model can also largely impact the performance of the sleep staging. In the earlier contributions, different classification algorithms deployed to perform the EEG sleep stages classification based on the Artificial Neural Network (ANN) [15], Support Vector Machine

Table 1
Changes sleep behaviour with respect to individual sleep stage.

Sleep Stage	Characteristics of the wave forms	Frequency Ranges(Hz.)
Wake(W)	$\alpha and \beta$	$8 \leq \alpha < 13 and 13 \leq \beta < 30$
NREM (N1)	$\alpha and \theta$	$5 \leq \theta < 7$
NREM (N2)	$\sigma and k-complex$	$12 \leq \sigma < 15 and 0.5 \leq k-complex < 1$
NREM (N3)	δ	$0.5 \leq \delta \leq 4$
REM	$\alpha, \beta, \theta, and Sawtooth$	$2 \leq sawtooth < 6$

Table 2
Correlation between wave patterns and age.

Age Range	Behaviour of sleep patterns
20–60	Wave patterns are slow with decrease of age. Similarly, the fast wave signal components are increased with the age of the subject
>60	Maximum slow wave forms are visible

(SVM) [11,28], linear discriminate analysis (LDA) [10], AdaBoost [28], K-Nearest Neighbor (KNN) [28], Naïve Bayes [28], Multilayer Perceptron [29], Random forest (RF) [17,29], Deep Learning (DL) techniques such as CNNs [30,31] and RNNs [32,33,34].

1.1. Related work

This section discusses related research studies available in the literature. Most of the proposed studies rely on EEG signals. These studies recommend the extraction of features from the representative input signals. Moreover, these studies also suggest using the different feature reduction techniques for identifying the suitable features. Finally, different classification algorithms have been used to classifying the sleep stages.

Oboyya et al. [35] proposed using single-channel EEG signals for sleep stage scoring in selected subjects between 35 and 50 years old. Moreover, they proposed wavelet transform techniques and a fuzzy c-means algorithm for sleep stage classification. The model achieved an overall accuracy of 85%.

In [36], the author proposed feature weighting method using K-means clustering. Welch spectral transform was used for feature extraction, and the selected features were used with K-means and decision tree techniques. The study reported an overall accuracy of 83%.

Aboalayon [37] also designed a sleep stage classification model based on EEG signals. The authors used a Butterworth bandpass filter to segment the EEG signals. These signals were then decomposed into different sub-bands, such as delta, theta, alpha, beta, and gamma. The extracted features were used with an SVM classifier. The work reported 90% classification accuracy.

The authors of [38] proposed the use of bootstrap aggregation for classification. This method was applied on single-channel EEG signals from two benchmark public sleep datasets, the Sleep-EDF and DREAMS datasets. The proposed system presented an accuracy of 92.43% for a two-state sleep stage classification problem.

In [39], the author proposed sleep stage classification based on time-domain features and structural graph similarity. The experimental work relied on single-channel EEG signals. The proposed SVM classifier presents an average classification accuracy of 95.93%.

Kristin M. Gunnarsdottir et al. [40] extracted both time and frequency domain properties from the PSG signal. In this study, the authors only considered healthy individuals with no prior history of sleep disorders. The extracted properties were classified using decision table classifiers. The reported overall accuracy was 80.70%.

Sriraam, N. et al. [41] used multi-channel EEG signals from ten healthy subjects in a proposed automatic sleep stage scoring model that considers sleep stage 1. In this study, spectral entropy features were extracted from input channels to identify irregularities in different sleep stages. A multi-layer perceptron feedforward neural network was used, and the overall accuracy of the model with 20 hidden units was reported as 92.9%. Moreover, using 40, 60, 80, and 100 hidden units, the proposed method reported 94.6, 97.2, 98.8, and 99.2% accuracy, respectively.

Memar, P. et al. [13] proposed a system to classify sleep and wake stages. The authors selected 25 suspected sleep disorder subjects and 20 healthy subjects for the experimental tests. 13 features were extracted from each of eight (alpha, theta, sigma, beta1, beta2, gamma1, and gamma2) sub-band epochs. The extracted features were validated using the Kruskal-Wallis test, then classified with a random forest classifier. The overall accuracy obtained from 5-fold cross-validation and subject-wise cross-validation was 95.31% and 86.64%, respectively. Da Silveira et al. [42] used discrete wavelet transform techniques for analyzing sleep behavior changes in different frequency ranges. The skewness, kurtosis, and variance features were extracted from the respective input channels. The random forest classifier that was tested for its ability to discriminate the various sleep stages showed an overall accuracy of 90%.

Xiaojin Li et al. [43] proposed a hybrid automatic sleep scoring system using EEG signals. Finally, the selected features were forwarded to random forest classifiers. The overall classification accuracy of the proposed method was 85.95%.

Zhu, G et al. [11] proposed sleep stage classification methods based on time and frequency domain features from single-channel EEG signals. The EEG signals were mapped onto visibility graphs and a horizontal graph to detect gait-related movements. Finally, nine features that were extracted from the input signals were forwarded to SVM classifiers considering multiple sleep stages. The proposed method presented an accuracy of 87.50% for two-state sleep stage classification problems.

Eduardo T. Braun et al. [44] proposed a portable sleep staging classification system using different features from EEG signals. The proposed method presented the best classification accuracy of 97.1% for the two-state sleep stage classification problem.

1.2. Contribution

This research proposes a new methodology in the automated sleep staging system, which is somehow dissimilarly incomparable to the existing methods discussed above. In this research, the ensemble learning stacking model is used to classify the sleep stages. This research work explores the effects of the age factor during the analysis of sleep behavior, which was ignored in many recent contributions. Still, it has been seen that there should be a direct association between sleep patterns and the subject age. This information helps to study the changes sleep characteristics in the different sleep stages The recent contribution of the sleep stages has been observed because many of these studies have ignored some important aspects, such as ignoring the age factor during the analysis of the sleep behavior, not addressing the unbalanced sleep epochs concerning the individual sleep stages. Also, it has been observed that infrequent sleep stages transitions occurred in many healthy category subjects, where the subjects are directly moved into a deep sleep from the wake stage [45,46]. Also, some of the existing studies have used only one classification algorithm for sleep scoring. All these above mentioned to be addressed in our proposed sleep staging study. In the proposed research work, we proposed an improved and efficient automated sleep scoring system based on a single-channel of EEG signals with improved classification accuracy incorporating with the existing studies.

The proposed sleep staging study has several advantages over the existing contribution. The first most advantage is proposed an ensemble learning stacking algorithm, which can effectively resolve the class imbalance problem in sleep staging. The second most advantage of this work is considered age as one of the features which helps to analyze the sleep behavior very effectively, which affects improving the classification accuracy.

This study was performed on two different datasets: the sleep EDF dataset and the sleep EDF expanded dataset in four steps. First, we considered all the extracted features for sleep staging with the inclusion and exclusion of age features using an ensemble learning stacking model. Second, the final selected top 29 features fed into the ensemble-based classifier for sleep staging. First, the extracted EEG recordings decomposed into several sub-bands of the signals then we applied the preprocessing techniques for eliminating the irrelevant noise compositions and muscle artifacts from the raw signals. After that, several features were extracted to analyze the changes in signal behavior concerning time and frequency levels. The second bagging approach was obtained to split the training data into multiple training sets as a sample subset. Third, the extracted features were screened based on weight value to generate the compact sample subsets, the compact subsets were forwarded into the base-layer classifiers to train the model proposed ensemble-based stacking model. Fourth, the testing data are forwarded into base-classifiers for obtaining the initial predictions. Finally, the base-layer predictions input into the meta-layer classification model for final predictions to complete the sleep staging classification.

To evaluate the effectiveness of the proposed methodology, several performance metrics obtained to compare with the other existing approaches. The proposed model's entire experiment executed under the R&K rules. We evaluated the proposed algorithm using 5-fold cross-validation techniques on two benchmark sleep EEG datasets, widely used in many of the existing contributions.

1.3. Structure of the paper

The entire research work is presented in this paper in six parts. In the first part, we introduce the importance of sleep study, describing some related contributions to sleep staging. The second part brief description of our proposed methodology includes descriptions of the database, signal preprocessing, feature extraction, feature selection, proposed ensemble learning stacking model, performance evaluation metrics, and testing schemes. The third part describes the complete experiment result analysis. The fourth part presents discussion and result comparisons between proposed performance results with existing similar published research works. The last part of this paper has a conclusion about this research work.

2. Materials and methods

This paper proposes an efficient and reliable automatic sleep staging classification system based on single-channel EEG signals using ensemble learning stacking algorithm techniques. Fig. 1 presents the steps of the proposed sleep staging system, and the following sub-stages have described the detail on each step. During feature extraction, both time and frequency domain features are extracted, and we extract one new feature as age from the enrolled subjects. We obtain the ReliefF weight feature selection algorithm for screening the relevant features from the

extracted feature vector during the feature selection step. Moreover, we also use the class balance strategy to address the unbalanced classification problem, and finally, we obtained an ensemble learning

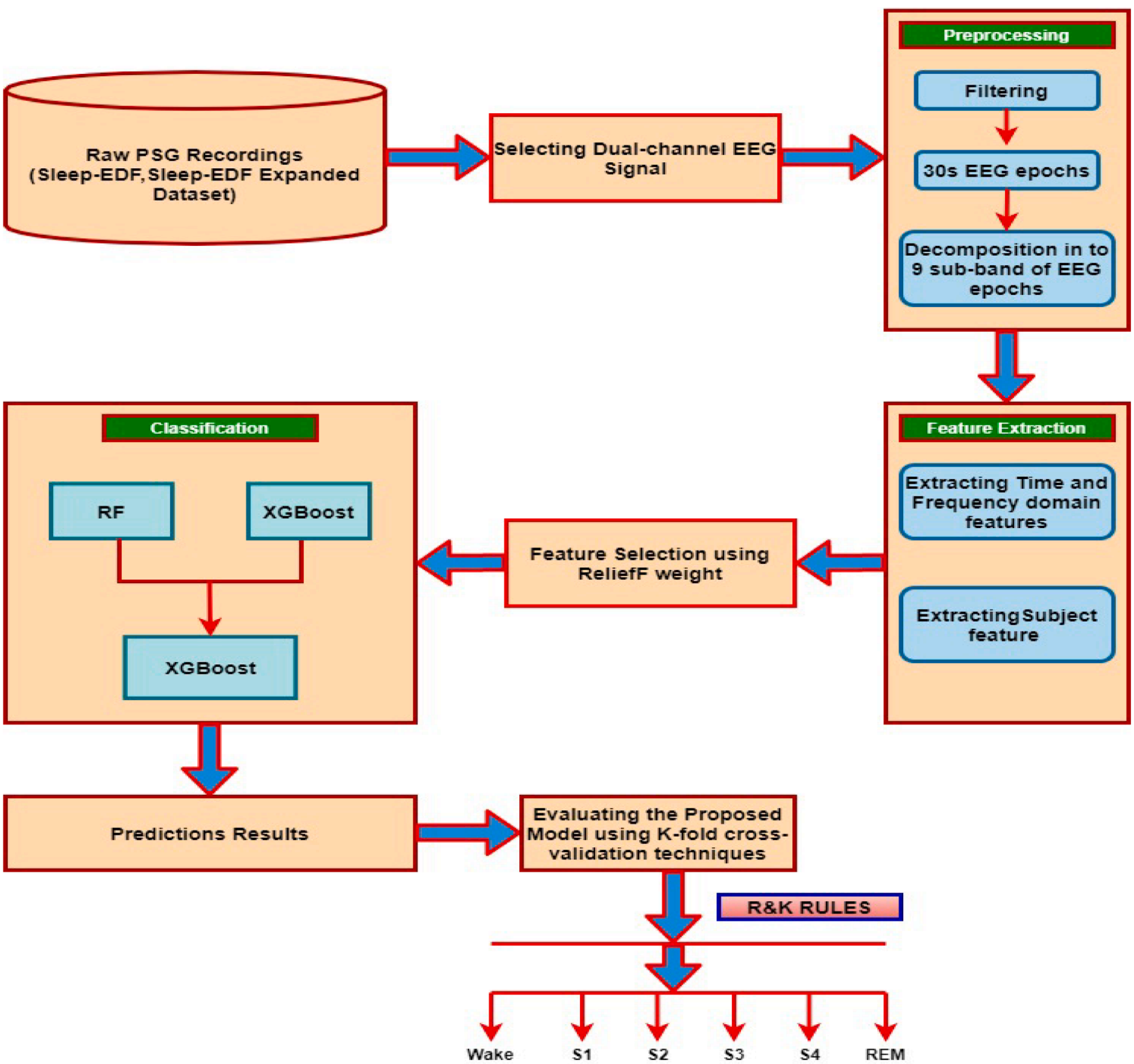


Fig. 1. Structural framework of the proposed automated sleep staging system.

stacking algorithm for improving the sleep staging classification performance.

2.1. Experimental data

This paper used two public sleep datasets, such as the sleep-EDF dataset(S-EDF), Sleep-EDF expanded dataset (SE-EDF) dataset for method evaluation. The records obtained from the two datasets are presented in Tables 3 and 4, respectively. Similarly, the details about the scoring procedures and sleep stages epochs are briefly mentioned in Table 5.The recorded sample sleep behavior from all the sleep stages under AASM sleep standards using S-EDF and SE-EDF databases are shown in Fig. 2.

2.1.1. Sleep-EDF (S-EDF) database

In this dataset, a total of 8 Caucasian subjects' full PSG recordings are collected with age ranges from 21 to 35, who were not taken any types of medication during the recordings. These recordings are grouped into two major subcategories, one from sc* and the other from st* section. The four subjects from sc* section are completely healthy controlled. The st* categories contained 4 subjects who were affected with mild difficulty during the sleep. For both category subjects, the recordings of 2 EEG signals (Fpz-Oz, Pz-Oz) and 1 EOG signal recorded. The details about the subject recordings are presented in Table 3. The sampling rate of the recorded signals is 100 Hz. The annotation of the sleep stages is done by sleep experts manually according to the R&K sleep study, which is labeled as Wake-W, NREM-N1, NREM-N2, NREM-N3, NREM-N4, and REM-R [47].

2.1.2. Sleep-EDF Expanded (SE-EDF) database

It is an extended version of the sleep EDF dataset, containing 61 EEG recordings from 20 healthy subjects and the subjects' age ranges from 21 to 34 years. All without any types of sleep medication [48]. Like Sleep-EDF recordings, these signals are also sampled with a sampling frequency of 100 Hz, and after that, the domain experts have labeled the sleep epochs based on the R&K sleep standards.

2.2. Pre-processing

The biomedical signals are recorded from subjects by placing the different electrodes in the different parts of the body, which are called raw signals. Generally, the raw signals can have different magnitudes. It is also associated with the sensor non-linearity, noises, and different types of artifacts like muscle twitching, motion, and eye blinks, limiting the analysis of the changes in sleep characteristics of the different sleep

Table 3
Recordings obtained from S-EDF dataset.

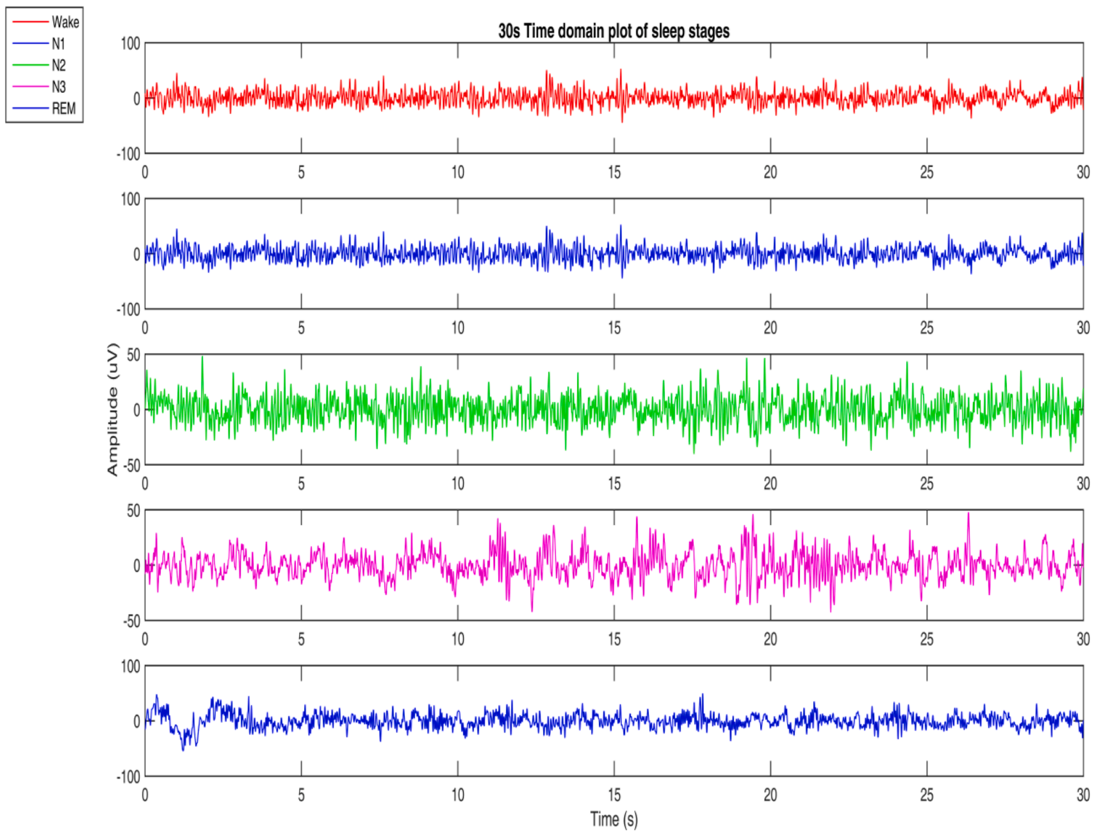
Sleep-EDF Database Sleep Cassette and Sleep Telemetry Study and Data			
Sl.No.	Name	Size	Modified
1	sc4002e0.hyp	6.0 KB	09-12-2002
2	sc4002e0.rec	49.2 MB	11-09-2007
3	sc4012e0.hyp	6.1 KB	09-12-2002
4	sc4012e0.rec	49.6 MB	11-09-2007
5	sc4102e0.hyp	6.1 KB	09-12-2002
6	sc4102e0.rec	49.7 MB	11-09-2007
7	sc4112e0.hyp	5.9 KB	09-12-2002
8	sc4112e0.rec	48.4 MB	11-09-2007
9	st7022j0.hyp	3.1 KB	09-12-2002
10	st7022j0.rec	31.1 MB	09-12-2002
11	st7052j0.hyp	3.0 KB	09-12-2002
12	st7052j0.rec	30.3 MB	09-12-2002
13	st7121j0.hyp	2.9 KB	09-12-2002
14	st7121j0.rec	29.2 MB	09-12-2002
15	st7132j0.hyp	2.8 KB	09-12-2002
16	st7132j0.rec	27.2 MB	09-12-2002

Table 4
Recordings obtained from SE-EDF dataset.

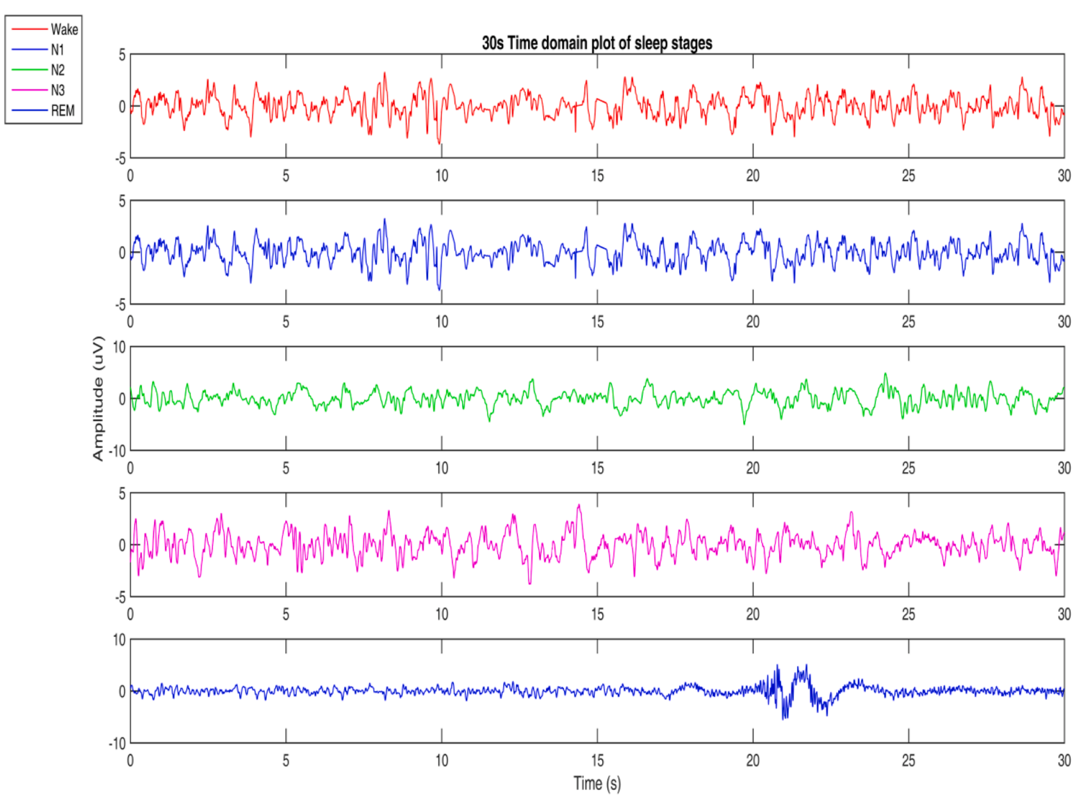
Sleep-EDF Database Expanded Sleep Cassette Study and Data			
Sl. No.	Name	Size	Modified
1	SC4001E0-PSG.edf	46.1 MB	05-04-2018
2	SC4001EC-Hypnogram.edf	4.5 KB	05-04-2018
3	SC4002E0-PSG.edf	49.2 MB	05-04-2018
4	SC4002EC-Hypnogram.edf	4.5 KB	05-04-2018
5	SC4011E0-PSG.edf	48.7 MB	05-04-2018
6	SC4011EH-Hypnogram.edf	3.8 KB	05–04-2018
7	SC4012E0-PSG.edf	49.6 MB	05-04-2018
8	SC4012EC-Hypnogram.edf	5.0 KB	05-04-2018
9	SC4021E0-PSG.edf	48.8 MB	05-04-2018
10	SC4021EH-Hypnogram.edf	4.7 KB	05-04-2018
11	SC4022E0-PSG.edf	47.9 MB	05-04-2018
12	SC4022EJ-Hypnogram.edf	5.2 KB	05-04-2018
13	SC4031E0-PSG.edf	49.1 MB	05-04-2018
14	SC4031EC-Hypnogram.edf	3.6 KB	05-04-2018
15	SC4032E0-PSG.edf	47.5 MB	05-04-2018
16	SC4032EP-Hypnogram.edf	3.7 KB	05-04-2018
17	SC4041E0-PSG.edf	44.7 MB	05-04-2018
18	SC4041EC-Hypnogram.edf	4.7 KB	05-04-2018
19	SC4042E0-PSG.edf	48.6 MB	05-04-2018
20	SC4042EC-Hypnogram.edf	5.2 KB	05-04-2018
21	SC4051E0-PSG.edf	47.4 MB	05-04-2018
22	SC4051EC-Hypnogram.edf	3.9 KB	05-04-2018
23	SC4052E0-PSG.edf	48.8 MB	05-04-2018
24	SC4052EC-Hypnogram.edf	4.1 KB	05-04-2018
25	SC4061E0-PSG.edf	48.2 MB	05-04-2018
26	SC4061EC-Hypnogram.edf	2.6 KB	05-04-2018
27	SC4062E0-PSG.edf	49.2 MB	05-04-2018
28	SC4062EC-Hypnogram.edf	3.1 KB	05-04-2018
29	SC4071E0-PSG.edf	48.9 MB	05-04-2018
30	SC4071EC-Hypnogram.edf	3.5 KB	05-04-2018
31	SC4072E0-PSG.edf	48.2 MB	05-04-2018
32	SC4072EH-Hypnogram.edf	5.2 KB	05-04-2018
33	SC4081E0-PSG.edf	48.6 MB	05-04-2018
34	SC4081EC-Hypnogram.edf	4.2 KB	05-04-2018
35	SC4082E0-PSG.edf	45.8 MB	05-04-2018
36	SC4082EP-Hypnogram.edf	4.6 KB	05-04-2018
37	SC4091E0-PSG.edf	47.5 MB	05-04-2018
38	SC4091EC-Hypnogram.edf	4.4 KB	05-04-2018
39	SC4092E0-PSG.edf	49.7 MB	05-04-2018
40	SC4092EC-Hypnogram.edf	3.6 KB	05-04-2018

Table 5
The detailed explanation of the sleep dataset obtained in this proposed research study.

Sleep scoring standards	Rechtschaffen and kales (R&K) manuals	
Database	Sleep-EDF *(S-EDF)* (Scored by one sleep expert)	Sleep-EDF (Expanded) *(SE-EDF)* (Scored by one sleep expert)
Number of Subjects	08	20
Gender (Male/ Female)	04/04	10/10
Age(years)	21–35	21–35
Epoch length (seconds)	30 s	30 s
Sampling Frequency(Hz.)	100 Hz	100 Hz
Total Sleep periods length(Hrs.)	~126.5	~342.5
EEG Montage	Bipolar	Bipolar
Channel	Fpz-Cz Pz-Oz	F_{PZ}-C_Z P_Z-O_Z
Sleep Stages		
Wake(W)	8006(53%)	6077(15%)
S1(N1)	604(4%)	2804(7%)
S2(N2)	3621(24%)	17779(44%)
S3(N3)	672(4%)	3370(8%)
S4	627(4%)	2333(6%)
REM	1609(11%)	7717(19%)
Total Epochs	15,139	40,100



(a)



(b)

Fig. 2. Sample EEG signals records obtained from the individual sleep stages: (a) using S-EDF database, (b) using SE-EDF database.

stages. It's highly essential to eliminate the irregular noises and artifacts to bring all the signals to the same scale. In this study for scaling and normalization, the classical zero mean and unit standard deviation method have been considered. By this method, the normalized i^{th} sample $(\overline{x}_i)$ of any signal$x$, with its mean μ and standard deviationσ, expressed as Eq. (1):

$$\overline{x}_i = \frac{(x_i - \mu)}{\sigma} \tag{1}$$

These artifacts, as mentioned earlier, were discarded using the 10th order Butterworth bandpass filter with a frequency range from 0.5 to 49.5 Hz in this study. These DC components are removed from the sleep recording using a lower passband filter with a frequency range of 0.5 Hz [16,49–53]. Similarly, the upper passband edge selection is made with a frequency range of 49.5 Hz. The original recorded signal and filtered signal for a single epoch are shown in Figs. 3 and 4. All the epochs are shown in Figs. 5 and 6 using S-EDF and SE-EDF databases respectively.

It has been found that recorded EEG signals are combinations of the different characteristics waveforms with different frequency ranges, which are highly correlated to different sleep stages. Henceforth it is more important to analyze the sleep behavior from the different characteristic waveforms with different frequency levels. It has been seen that it should be essential to obtain filtering approaches for the biomedical signals, which helps remove the baseline removal, denoising, smoothing, or sharpening. Most of the time, digital filters are more preferable rather than analog filters for flexibility. Filters are mainly extracting the desired signal information from the irrelevant and undesired noises and artifacts. Generally, the biomedical signals are of low amplitude, and most of the electrodes are having a high impedance. Therefore, these signals are highly acquiring the surrounding noises, and also sometimes, coupling of the other biomedical signals acts as one of the major disturbances. In this work, for processing the EEG signals, Finite Impulse Filters (FIR) are implemented. Its implementation is easy and flexible, and it is less computationally intensive in comparison to the other filters.

A discrete time-series based system is one of the mathematical models in which a discrete-time input signal $x[n]$ into an output signal $y[n]$.The mathematical form of the digital filter as follows:

$$y[n] = \sum_{k=1}^{K} a_k y[n-k] + \sum_{m=0}^{M} b_m x[n-m] \tag{2}$$

This mathematical equation helps to compute the $y[n]$ at time n from a finite number of input values $x[n]$ and the output.

Where$M$ = Maximum of Numbers

K=Order of the filter

The digital filters can be classified based on the coefficient values mentioned in Eq. (2).If all the coefficients a_k being zero in Eq. (3) then the output calculated based on the finite numbers of the input values. So in this condition, such filters are called Finite Impulse Response (FIR) (or) moving average (MA) filters. The term MA means it holds the out of the average samples of inputs. The most general mathematical form of moving average filter for averaging over N consecutive epochs of length L is formulated as follows:

$$y[n] = \frac{1}{N} \sum_{k=0}^{N-1} x[n-kL] \tag{3}$$

In this study, the recorded EEG signals are segmented into different frequency bands with different frequency levels to discriminate the sleep stages better. We divided into two different waveforms for better discrimination between the wake and REM sleep stages, such as β1 (13–20 Hz), and β2 (20–30 Hz). Similarly, from the study [13], it has been reported that the waveforms have a significant impact with regards to analysis the changes in sleep characteristics in the individual sleep stages so that we divided them into γ1 (30–40 Hz) and γ2(40–49.5 Hz) waveforms. This proposed study's different waveform boundary is defined as follows: δ wave: 0.5–4 Hz, θ wave: 4–8 Hz, σ wave: 12–15 Hz, and k-complex wave: 0.5–1 Hz. Finally, nine Finite Impulse Response (FIR) bandpass filters are obtained to separate the δ, θ, α, σ, β1, β2, γ1, γ2, and k-complex wave characteristics with frequency ranges of 0.5–4 Hz, 4–8 Hz,9–11 Hz,12–15 Hz,14–20 Hz,20–30 Hz, 30–40 Hz,40–49.5 Hz, and 0.5–1 Hz respectively. The changes in sleep behavior of the EEG

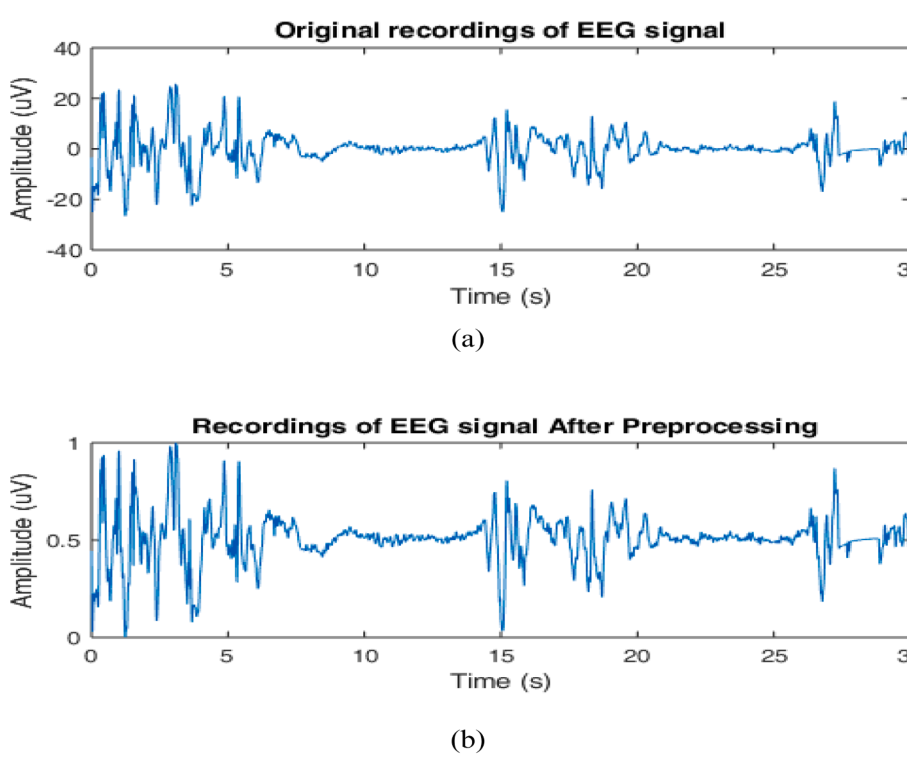


Fig. 3. Comparison of the signal-epoch obtained from the S-EDF database: (a) the original signal (b) the filtered signal.

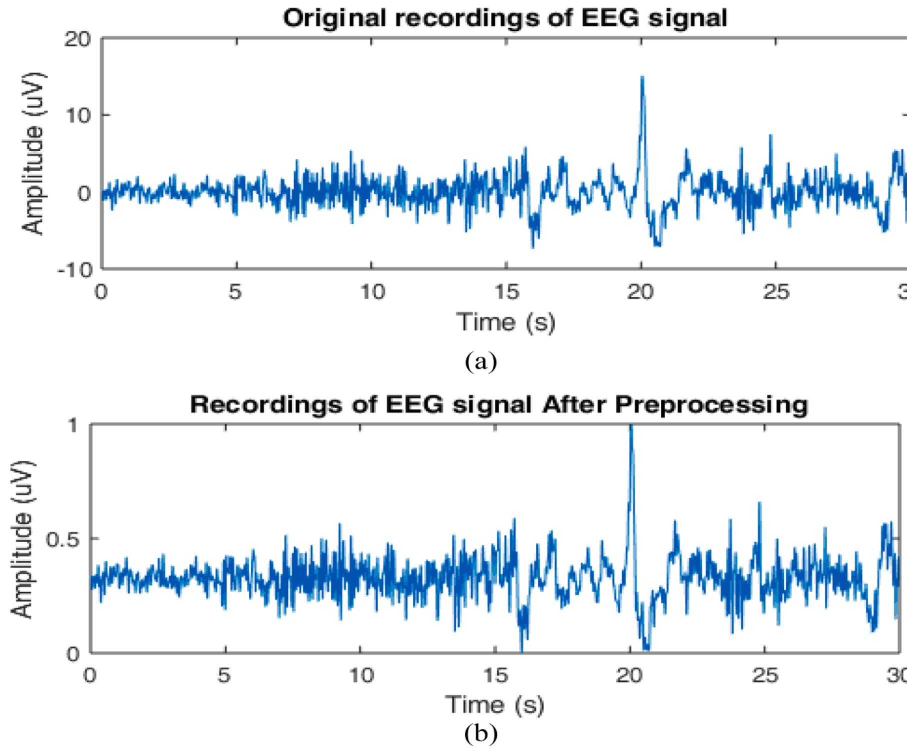


Fig. 4. Comparison of the signal epoch obtained from the SE-EDF database: (a) the original signal (b) the filtered signal.

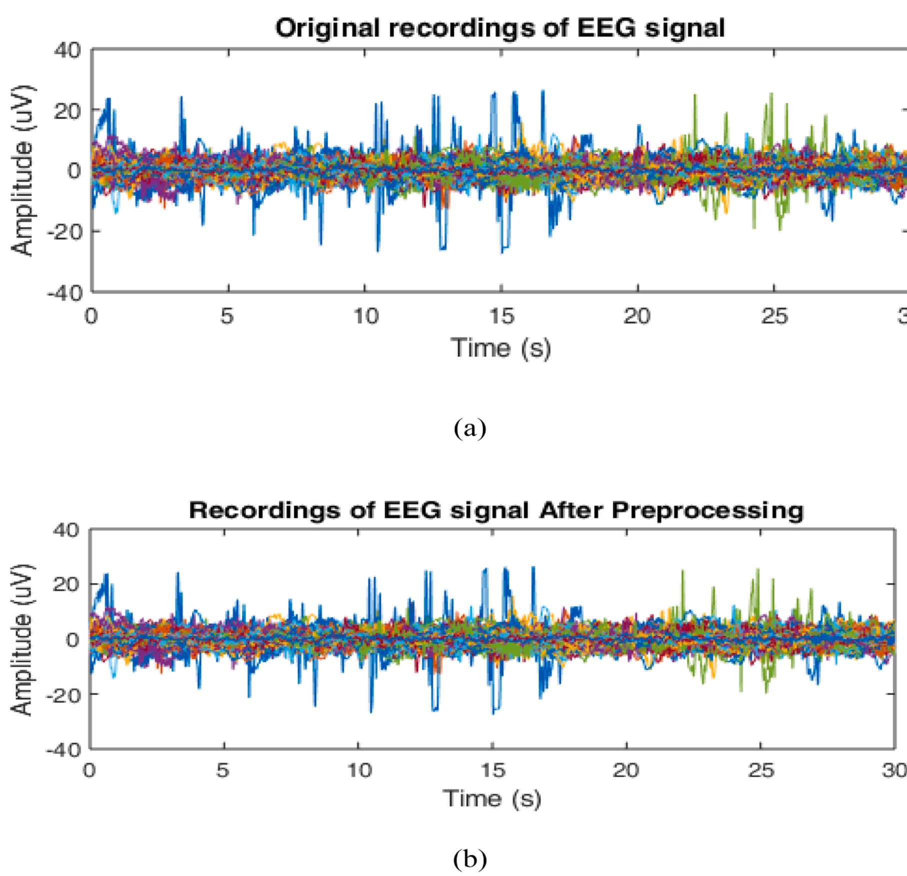


Fig. 5. Comparison of all the epochs obtained from the S-EDF database: (a) the original signal (b) the filtered signal.

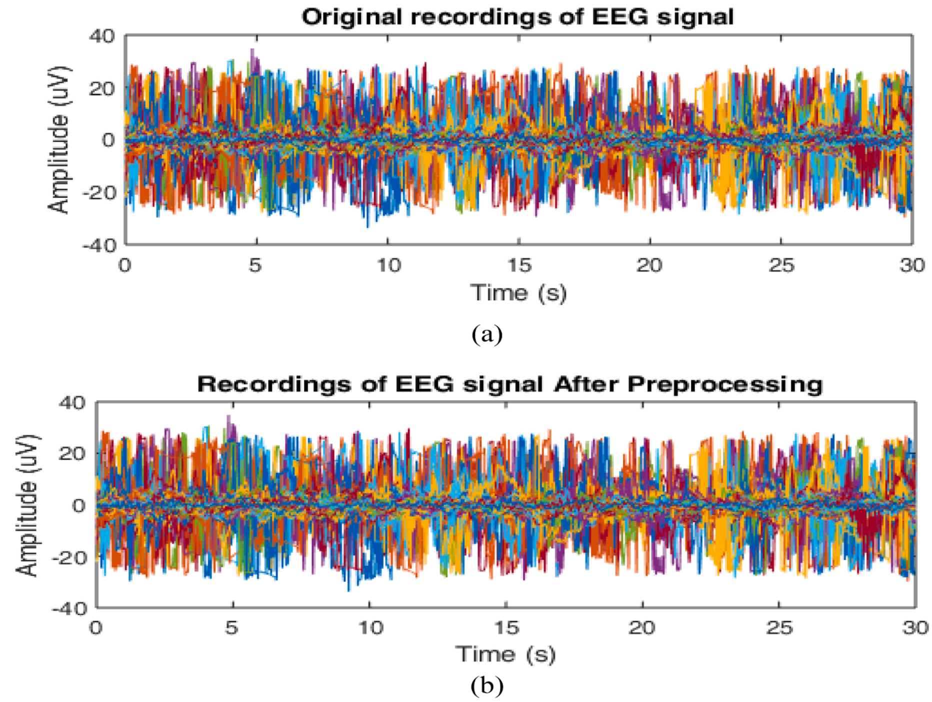


Fig. 6. Comparison of all the epochs obtained from the S-EDF database: (a) the original signal (b) the filtered signal.

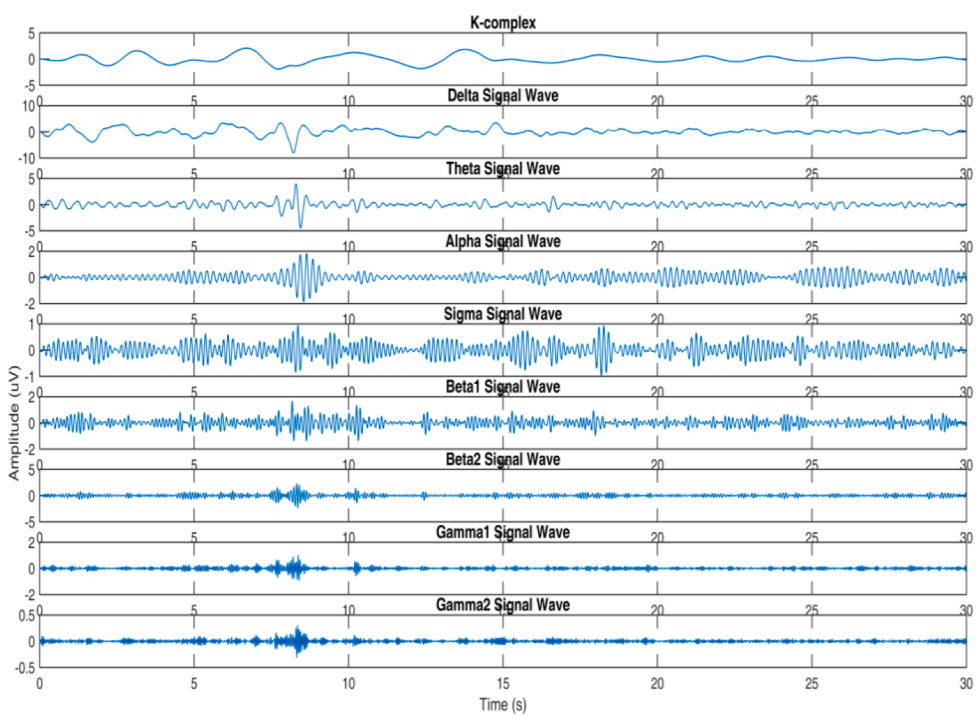


Fig. 7. The changes characteristics obtained of the EEG signal sub-band using FIR band-pass filters.

signal sub-bands are shown in Fig. 7.

2.3. Feature extraction

Feature extraction is an important step for automated analysis of subjects' sleep behavior from the EEG signals, and feature-based analysis has been most effective for identifying the subjects' different sleep characteristics. It has been observed from further sleep studies that proper analysis of features may easier during sleep stages classification because it directly impacts the classification performances. As we know, brain EEG signals are highly random and non-stationary. For that reason, extracting the most relevant features is important for properly interpreting the sleep stages' behavior. Most of the contributions majorly focused on the three different feature categories for signal analysis in the recent research progress, such as time-domain, frequency-domain, and non-linear features [54–68]. In this work, we obtain both time and frequency domain features. The time-domain analysis helps to retrieve information of fluctuating sequences of the signal and detects the epileptiform discharges. It's very important for extracting the frequency domain features for sleep scoring to discriminate the sleep characteristics related to the individual sleep stages from EEG signal because it provides the changes of delta (δ) rhythm, theta (θ) rhythm, alpha (α) rhythm, sigma (σ) rhythm, and beta1 (β1) rhythm, beta2 (β2) rhythm, gamma1 (γ1) rhythm, gamma2 (γ2) rhythm and k-complex patterns in the EEG signals in a different frequency range. In this proposed research work, we extracted 28 features, which include 1.) 13 time-domain-based features, 2.) 15 frequency domain-based features.

2.3.1. Statistical parameters measures

In the proposed sleep staging study, we obtained the statistical approach to extract the time domain features from the input signal. Various statistical parameters were used for measuring the features with the correlated sub-bands. The statistical parameters: mean, standard deviation, minimum, maximum, median, variance, skewness, and kurtosis are considered to derivate the sleep behavior of the patients related to the individual sleep stages. It has been observed that among the various statistical parameters of EEG signal, the first-to-fourth order moments such as mean (1st raw moment M1), variance (2nd central moment M2), skewness (normalized 3rd central moment M3), and kurtosis (normalized 4th central moment M4) were computed from the 30 s epochs of the EEG signal to measure the central tendency, degree of dispersion, asymmetry, and peakedness respectively. The variance (M2) helps to interpret the sleep behavior in the REM sleep stages from the NREM N2 and N3 stages. The third quartile (Q3) helps find the values below 75% of the random variable values are identified. This feature was also obtained in the different existing studies for discriminating the characteristics of sleep stages. Zero crossing rate (ZCR) provides count information regarding the number of times the EEG signal crosses the relevance line obtained from the mean value. This feature is quite suitable for the characterization of the sleep spindles, and it also helps to analyze the sleep stages activities from the EEG signal.

2.3.2. Hjorth parameters

It is one of the popular parameters among the time-domain features for analysis of the EEG signals and is also used for sleep scoring. It provides dynamic temporal information of EEG signals, which helps interpret sleep behavior during sleep. The Hjorth parameters are calculated to derive the features indicative of activity, mobility, and complexity. All these three parameters are derived based on the variance of the derivatives of the EEG signal waveform. The variations of the Hjorth Parameters at different sleep stages are presented in Fig. 8.

2.3.3. 75th percentile

It helps find the values below 75% of the random variable values are identified. This feature was also obtained in the different existing studies for discriminating the characteristics of sleep stages.

2.3.4. Zero crossing rate

It provides count information regarding the number of times the EEG signal crosses the relevance line obtained from the mean value. This feature is quite suitable for the characterization of the sleep spindles, and it also helps to analyze the sleep stages' activities from the EEG signal.

Typical examples of the EEG signal's sleep stage measured through the EEG channel of the PSG system with a sampling frequency of 100 Hz and the 200 points moving average are presented in Fig. 9.

2.3.5. Frequency domain feature analysis

EEG signal has a strong background of frequency-domain characteristics, and it plays an important role during sleep staging to characterize the different sleep EEG rhythmic waveforms. Generally, frequency-related features were better present in the lower frequencies of the EEG signal during the depth of the sleep [16]. It has been found that the spectral features are mostly used for monitoring the changes in the behaviour of the sleep with regards to the different sleep stages. For example, alpha and beta wave patterns have mostly appeared during Wake and REM sleep periods, and the delta and theta patterns are generally seen in the NREM N3 stage.

2.3.5.1. Spectral power in the different frequency sub-bands. It's important that first of all, we compute the spectral power from the delta (δ) band with 0.5–4 Hz, theta (θ) band with 4 to 8 Hz, alpha (α) band with 8 to 12 Hz, and beta (β) band from 13 to 30 Hz of the EEG signals. Further, it helps to compute the other important features like relative spectral power, power ratios, and power spectrum from an epoch of the EEG signal.

2.3.5.2. Relative spectral power. This feature computed for the frequency sub-bands (δ, θ, α, and β) of the EEG signals. The ratio was computed by considering the average power ratio between each specific range and the total power, where the total power was the summation of the individual bands [17]. The variation of relative spectral power in the four frequency bands and different sleep stages is presented in Fig. 10.

2.3.5.3. Power ratios. The power ratios are computed in between the relative spectral powers in different frequency bands.

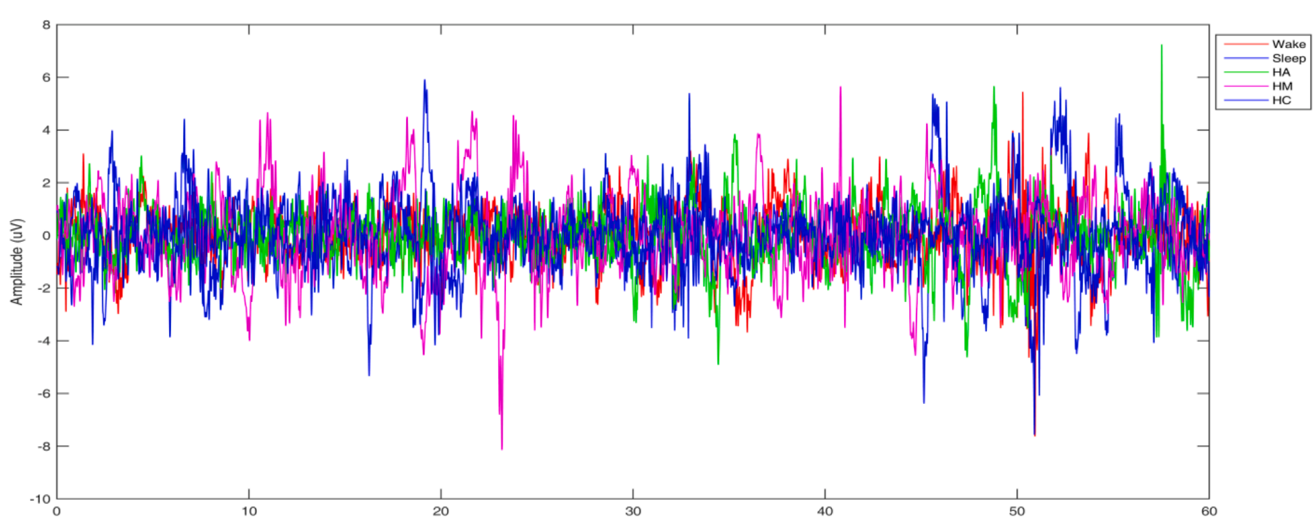


Fig. 8. Variations of Hjorth Parameters at different sleep stages.

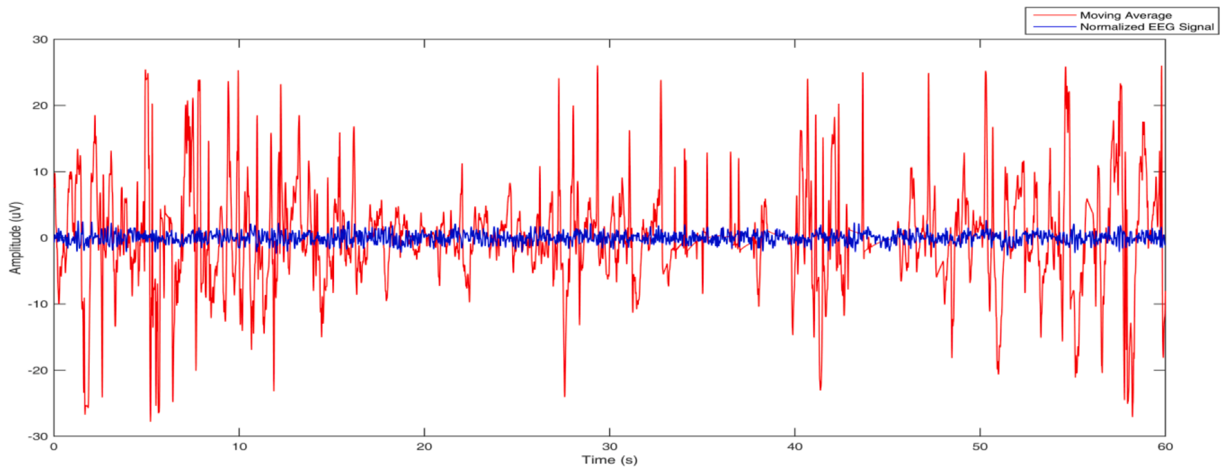


Fig. 9. EEG signal during sleep stage and 200 points moving average.

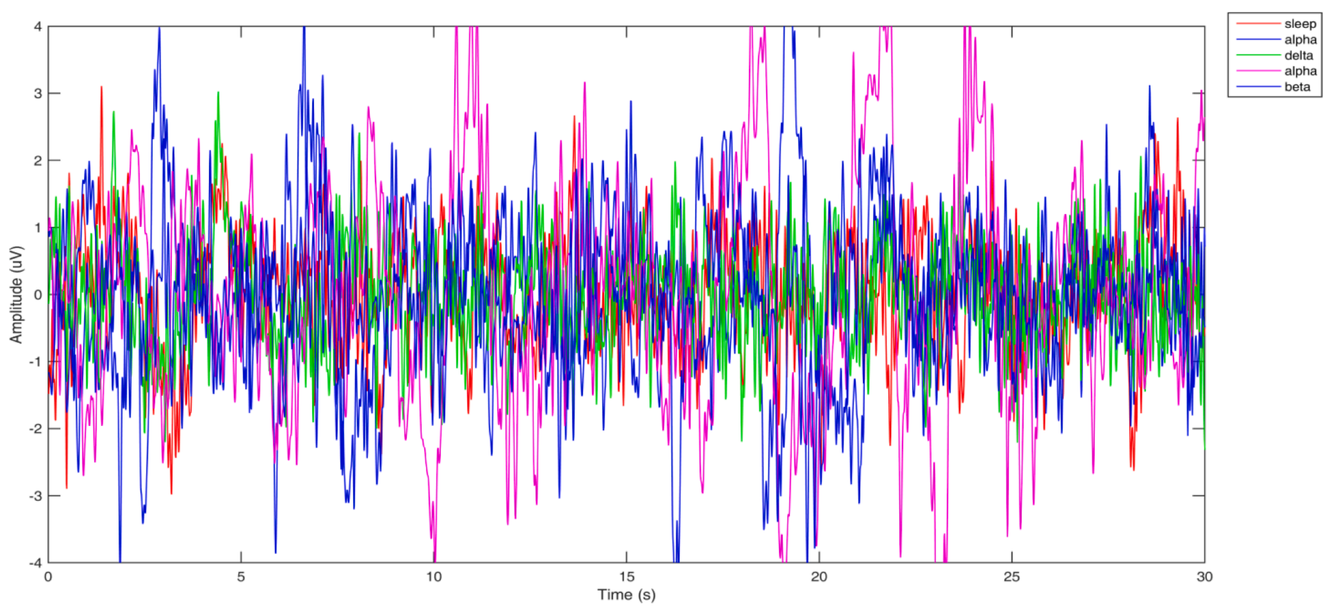


Fig. 10. Variation of relative spectral power at different sleep stages in the delta, theta, alpha, and beta frequency bands.

2.3.5.4. Power spectrum. This feature describes the distributions of the power into different frequency bands. This feature helps to monitor the changes in characteristics of the sleep in the different sleep periods. The changes in the power spectrum during the sleep stages are presented in Fig. 11.

The extracted different features in this proposed study and their descriptions are summarized in Table 6, were performed, and epochs of the 30-s length of the filtered EEG signals.

Algorithm 1: Pseudo-code for normalization of feature vector

Input: FV [1: N1, 1: M1]
Feature vector (N1: total number of sample points, M: total number of features)
Output: NFV [1: N1, 1: M1] Normalized Feature Vector
mean () and std () function used for computing mean and standard deviation values of the feature vector
Step 1: Create a vector with the name NFV with dimension size NXM and create two other vectors named MV [1, 1: M1] and SV [1, 1: M1]
Step 2: MV [1, 1: M1] ←mean ()
Step 3: SV [1, 1: M1] ←std ()
Step 4: for i←1 to M1, do
Step 5:NFV[1:N1,i]←(FV[1:N1,i]−MV[1,i])/SV [1,i]
Step 6: end for

2.4. Feature screening

This step aims to select the most relevant features from the pool of the extracted features. It has been seen several times that all extracted features may not be proper for all the subject cases for discriminating the sleep patterns. It may be one of the causes for the degradation of the classification results. In this research work, we have used the ReliefF feature selection algorithm for identifying suitable features. It is one of the supervised feature weighting algorithms that evaluate the

features' relevance concerning its class labels [69]. This algorithm's essential concept is to select highly commendable features that help discriminate the subject's sleep behavior. As an output, this algorithm assigned a weight to individual input features according to their relevance. It generates a weight for each feature, and the larger the feature's importance, the higher the association between the features and sleep stages. The main advantage of this algorithm is well managed with noisy and unknown data. The pseudo-code for the ReliefF feature selection algorithm is presented in Algorithm 2.

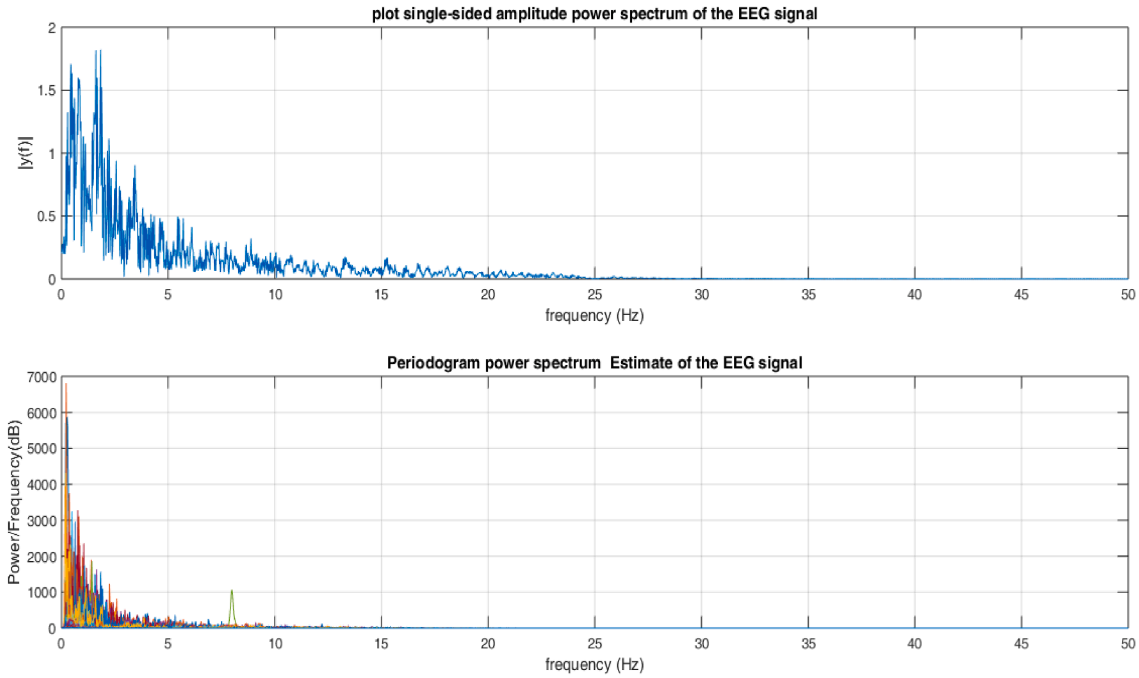


Fig. 11. Variation of power spectrum at different sleep stages obtained from the single-epoch and all-epochs.

Algorithm 2: Pseudo-code of the ReliefF feature selection algorithm

Require Input: Feature Vector with Sample values and its target class labels
Ensure Output: Weight vector *W [V]* contained the feature vector's features' weight corresponding features.
Step 1: Initially set all weights for features as zero *W [V]*:=0
Step 2: for i: =1 to r do start
Step 3: randomly choose the sample values of *Fi*
Step 4: find the k nearest hit points *Hj*(C)
Step 5: for each class C $\neq$ class (*Fi*) do
Step 6: find the k nearest missing points *Mj*(C) from class
Step 7: for *V*: =1 to a do
Step 8: *W [V]* = *W [V]* − *h1* + *h2*
Step 9: end

The algorithm starts with initializing the weight vector *W [V]*, initially for all the features, and the weight values are set as zero in the weight vector. We get the weight values for the corresponding features from the number of iterations. *r* denoted the number of iterations to be conducted. *Fi* is the sample features randomly selected from the feature spaces *I1, I2, I3*……*In,* and the sample class label denoted as C. Class(*Fi*) is indicated as the class label of *Fi*. H represents the k sample (hit) points that share the same class with *Fi*, similarly the k nearest missing points *M (C)* with class C, which is different from the class label of *Fi*. The k value denotes the user-defined nearest neighbor points.

$$h1 = \sum_{j=1}^{k} \frac{diff(V, F_I, H_J)}{rXk} \tag{4}$$

$$h2 = \frac{\sum_{C \neq Class(F_i)} \left(\frac{P(C)}{1-P(Class(F_i))} \right) \sum_{j=1}^{k} diff(V, F_I, M_J(C))}{(rXk)} \tag{5}$$

$$diff(V, I1, I2) = \frac{|value(V, I1) - value(V, I2)|}{\max(V) - \min(V)} \tag{6}$$

In Eqs. (4)–(6), P(C) contained information with related to misses weighted value with the prior probability of that class and 1 − P (Class (Fi)) presents sum of the probabilities non-Fi sample points;diff (V, I1, I2) presents calculation of distance difference results from sample points I1 to I2;value (V, I) presents the feature value of the particular instance I1.

The features and their corresponding ReliefF weight values complete lists for S-EDF, and SE-EDF datasets are presented in Tables 7 and 8 respectively.

Through the ReliefF weight feature selection algorithm, we screen the relevant features from the extracted feature vectors, but it cannot discriminate the redundant features; majorly, the redundancy occurred due to the strong correlation between the features. Eliminating those redundant features may increase the sleep staging performance. In this proposed sleep staging study, the Pearson correlation coefficient was obtained to find the correlation between the features. The mathematical form of computation of the Pearson correlation coefficients is shown in Eq. (7).

$$\rho(X, Y) = \frac{E[(X - \mu_x)(Y - \mu_Y)]}{\sigma_x X \sigma_y} \tag{7}$$

where μ the average is value and σ is the standard deviation. E is the mathematical expectation. Those Pearson coefficients are considered highly relevant, whose absolute value is greater than or equal to the threshold value of 0.90. Those features to be discarded, which are having low weights. We have considered the first 16 features for the S-EDF dataset and the first 18 features for the SE-EDF dataset. Further, we have applied Pearson correlation analysis upon these selected features for eliminating redundant features. The calculated Pearson coefficients matrix for S-EDF and SE-EDF are presented in Tables 9 and 10, respectively

The order number in Table 9 is the same as the weight sorting result in Table 7. From Table 9, it has observed that only two Pearson correlation coefficients feature pairs (i) pow_ratio2 and pow_ratio7, (ii) pow_ratio1 and RSP_alpha which threshold values are greater than 0.90. These two pairs of redundant features are removed from 16 selected features and rest of 14 features are selected for further classification task.

From the SSE-EDF dataset, only the first 18 features are relevant according to the ReliefF weights of the features. The features and their corresponding ReliefF weights are described in Table 8. We have computed a correlation matrix for identifying the redundant features, and the order is maintained in Table 10 is the same as Table 8, which

Table 6
The short explanation of the extracted features for this proposed study.

Feature No.	Extracted Feature Equation	Feature Descriptions
Time-domain based features		
1	$Mean(\overline{x}) = \frac{1}{N}\sum_{i=1}^{N} x_i$ with $N =$ the length of the data sample x and $\overline{x}$ is the mean of the data sample	The mean electrical potential of an epoch is calculated. It also measures the central tendency in the data points.
2	$Maximum = \mathrm{Max}[x_i]$	It is used to quantify the range of data and find the magnitude of the signal baseline.
3	$Minimum = \min[x_i]$	
4	$Variance(Var) = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}$	It helps to determine how the data is dispersion concerning the value of the mean.
5	$Standrad\ deviation(SD) = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2\right)^2$ with N = the length of the data sample x and $\overline{x}$ is the mean of the data sample	It is used to calculate the quantity of variation and dispersion of the data.
6	$Median = \left(\frac{N+1}{2}\right)^{th}$ with N = the length of the data sample x	It helps to get the information about the center and spread of the signal data
7	$75^{th} percentile(75_P^{th}) = Max(x_i) < P\{75\}$ $x_i$is the signal data $p\{75\} =$ the 75th percentile of the signal data	The percentile analysis provides amplitude information of the signal, which helps to discriminate the sleep stages. It defines the value below which 75% of the random variables values data is located
8	$Signal\ Skewness(skew) = \sum_{i=1}^{N}\frac{\left(E(x_i - \overline{x})^3\right)}{\sigma^3}$ Where $N$is the length of the signal data x_i σ is the standard deviation of the sample data for all $i$$E$is the expected mean value $E(x) = \sum_{i=1}^{N} p_i x_i$ p_i presents the probabilities with associated to the signal data x_i	The skewness helps to measure the symmetry of the distribution of the signal concerning the mean value. The signal's normal distribution is zero, while the positive and negative skewness indicates that the data are skewed into the right and left-hand sides. It is a higher-order-statistics measure(third moment)
9	$Signal\ kurtosis(kurt) = \sum_{i=1}^{N}\frac{\left(E(x_i - \overline{x})^4\right)}{\sigma^4}$ Where $N$is the length of the signal data x_i σ is the standard deviation of the sample data for all i, and E is the expected mean value	It measures whether the data is peaked or flat relative to the normal distribution. It is a higher order-statistics measure (fourth moment)
10	$Signal\ Activity = Var(x_i)$	Among the time-domain features, Hjorth parameters (Activity, Mobility, and Complexity) are more popular for interpreting the EEG signals, directly helping during the classification of sleep stages. It provides the dynamic temporal information of the polysomnography signals. These parameters are computed based on the variance of the derivatives of the signal data.
11	$Signal\ Mobility = \sqrt{Var(x_i')/Var(x_i)}$	
12	$Signal\ Complexity = \sqrt{Var(x_i'')XVar(x_i)/Var(x_i')^2}$	
13	$Zerocrossing\ Rate(ZC) = \sum_{i=1}^{N} ZC'(i)$ $ZC'(i) = \begin{bmatrix} 1, x(i) \le 0 \cap x(i+1) \\ 1, x(i) \ge 0 \cap x(i+1) \\ 0, otherwise \end{bmatrix}$	It provides information with regards to the number of instances where the EEG signal crosses the references line.
Frequency-domain based features		
14,15,16,17, 18,19,20,21,22	$Relative\ Spectral\ Power(\delta,\theta,\alpha,\beta1,\beta2,\gamma1,\gamma2,\sigma,k-complex) = \frac{\int_{-f1}^{f0} x_i(f)df + \int_{f0}^{f1} x_i(f)df}{\int_{-\infty}^{\infty} x_i(f)df}$	It is one of the popular parameters among in the frequency domain features, which helps to compute the changes behavior in the different stages of the signal waveform from $\delta, \theta, \alpha, and\beta$ frequency sub-bands.
23	$Power\ Ratio1 = \frac{\delta}{\beta}$	It is used for computing the power from the epochs' current and background with the different frequency ranges for interpreting the sleep stages' behavior.
24	$Power\ Ratio2 = \frac{\delta}{\theta}$	
25	$Power\ Ratio3 = \frac{\theta}{\alpha}$	
6	$Power\ Ratio4 = \frac{\theta}{\beta}$	
27	$Power\ Ratio5 = \frac{\alpha}{\beta}$	
28	$Power\ Ratio6 = \frac{\alpha}{\delta}$	
29	$Power\ Ratio7 = \frac{\delta}{\alpha}$	
30	$Power\ Ratio8 = \frac{\delta}{\sigma}$	
31	$Power\ Ratio9 = \frac{\delta}{\gamma1}$	
32	$Power\ Ratio10 = \frac{\theta}{\sigma}$	
33	$Power\ Ratio11 = \frac{\theta}{\gamma2}$	
34	$Power\ Ratio12 = \frac{\alpha}{\sigma}$	
35	$Power\ Ratio14 = \frac{\alpha}{\gamma1}$	
36	$Power\ Ratio15 = \frac{\sigma}{\beta1}$	
37	$Power\ Ratio16 = \frac{\sigma}{\gamma1}$	
38	$Power\ Ratio17 = \frac{\beta1}{\gamma2}$	
39	$Power\ Ratio18 = \frac{\delta}{(\theta+\alpha)}$	
40	$\mathrm{Power\ spectrum}(\delta) = \sum_{i=1}^{N} x_{i\delta}$	It helps to retrieve how the intensity of time-series signal data is distributed in the frequency domain.
41	$\mathrm{Power\ spectrum}(\theta) = \sum_{i=1}^{N} x_{i\theta}$	
42	$\mathrm{Power\ spectrum}(\alpha) = \sum_{i=1}^{N} x_{i\alpha}$	

(continued on next page)

Table 6 (*continued*)

Time-domain based features		
Feature No.	Extracted Feature Equation	Feature Descriptions
43	Power spectrum$(\sigma) = \sum_{i=1}^{N} x_{i\sigma}$	
44	Power spectrum$(\beta 1) = \sum_{i=1}^{N} x_{i\beta 1}$	
45	Power spectrum$(\beta 2) = \sum_{i=1}^{N} x_{i\beta 2}$	
46	Power spectrum$(\gamma 1) = \sum_{i=1}^{N} x_{i\gamma 1}$	
47	Power spectrum$(\gamma 2) = \sum_{i=1}^{N} x_{i\gamma 2}$	
48	Power spectrum$(k_{comp}) = \sum_{i=1}^{N} x_{i\,k-complex}$	
49	Age	

Table 7
ReliefF weights with S-EDF dataset for 28 features.

Order	Feature number	Feature name	ReliefF weight
1	8	RSP_beta1	0.97
2	11	Signal mobility	0.86
3	16	RSP_delta	0.74
4	28	Pow_Ratio6	0.64
5	24	Pow_Ratio2	0.56
6	23	Pow_Ratio1	0.50
7	40	delta_powbp	0.46
8	44	Beta1_powbp	0.44
9	27	Pow_Ratio5	0.44
10	29	Pow_Ratio7	0.40
11	18	RSP_alpha	0.39
12	17	RSP_theta	0.38
13	42	alpha_powbp	0.12
14	25	Pow_Ratio3	0.09
15	41	theta_powbp	0.07
16	5	SD	0.04
17	10	Signal activity	0.04
18	4	Variance	0.04
19	12	Signal complexity	0.03
20	7	Percentile	0.02
21	8	Skew	0.01
22	25	ZC	0.01
23	3	Minimum	0.01
24	9	Kurt	0.01
25	6	Median	0.01
26	1	Mean	0.01
27	26	Pow_Ratio4	0.01
28	20	Maximum	0.01

Table 8
ReliefF weights with SE-EDF dataset for 28 features.

Order	Feature number	Feature name	ReliefF weight
1	17	RSP_theta	1.00
2	16	RSP_delta	0.95
3	18	RSP_alpha	0.78
4	23	Pow_Ratio1	0.77
5	11	Signal mobility	0.77
6	25	Pow_Ratio3	0.74
7	12	Signal complexity	0.73
8	41	theta_powbp	0.70
9	19	RSP_beta1	0.69
10	29	Pow_Ratio7	0.65
11	24	Pow_Ratio2	0.64
12	40	delta_powbp	0.56
13	42	alpha_powbp	0.34
14	44	Beta1_powbp	0.52
15	27	Pow_Ratio5	0.34
16	26	Pow_Ratio4	0.31
17	10	Signal activity	0.24
18	4	VAR	0.24
19	5	SD	0.23
20	1	Mean	0.14
21	6	Median	0.14
22	9	Kurt	0.09
23	2	Maximum	0.08
24	28	Pow_Ratio6	0.08
25	7	75th Percentile	0.06
26	8	Skew	0.04
27	3	Minimum	0.02
28	6	Median	0.01

presented ReliefF weights for the SE-EDF dataset. It has been observed that one pair of redundant features, i., variance and signal activity, found, which threshold value is greater than 0.90.Finally, we removed the one pair redundant features from 18 relevant features.

2.5. Proposed ensemble learning stacking model

Ensemble techniques are one of the machine learning approaches for improving the model by combining several models. The main important advantage of ensemble learning is to reduce the variance and bias factor. It helps to increase the accuracy of the model and reduce the variability of the prediction. Ensemble techniques can be approached in two ways such as sequential and parallel manner. In the sequential methods, the model considered the previous model performance in the sequence manner to improve the system's performance. This approach encompasses most machine learning models such as AdaBoost, XGBoost [70], etc. In the other part, the parallel approach, the models are independently performed, which helps learn the different forms of data learned better. This approach includes machine learning models such as Random Forest, Bagging, and Boosting. A different ensembling method is called ensembling learning of the stacking model, which is in two-layer architecture. The first layer models and data are considered base-layer data and models, respectively.

Similarly, the second layer of cross-validated data and models are named *meta*-classifiers. Our proposed model supports the parallel structure since the base-layers models are trained independently from each other. Here, the base-layers models' predictions become the meta layers inputs, and finally, the *meta*-classifier delivers the final predictions. The basic schema structure of the proposed ensembling stacking model is shown in Fig. 12.

The major advantage of the proposed stacking over the alternatives parallel ensembling approach can improve the *meta*-model analysis. It also helps to identify the miss-predicts samples continuously from the base-layers classifiers' feature space due to inconsistent learning. The same problem is easily recognized through the *meta*-model and suitable base-models suited for that specific feature space. The same challenges may not be possible for recognizing through the voting or simple aggregation of bagging approaches. Additionally, we also obtained cross-validation techniques to overcome the overfitting problem. In this work, we have used a 5-fold cross-validation approach; during this approach, the training dataset is split randomly into five equal-size folds, out of that 5-folds are used to fit for the training, the remaining 1-fold are treated for testing purpose.

This process is repeated five times for training the data using base-layer classification models Random Forest (RF), eXtreme Gradient Boosting (XGBoosting). The main parameters of the RF algorithm are the number of trees (m), the maximum number of iterations for the weak learner (n_estimators), and oob-score. The study parameters are set as

Table 9
Pearson correlation matrix between first 16 features.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.00	0.91	−0.80	−0.50	−0.47	0.62	−0.47	0.77	−0.63	−0.52	0.60	0.11	0.42	−0.42	0.04	−0.14
2	0.91	1.00	−0.88	−0.45	−0.61	0.66	−0.43	0.82	−0.65	−0.66	0.71	0.32	0.61	−0.58	0.27	−0.04
3	−0.80	−0.88	1.00	0.27	0.63	−0.79	0.56	−0.69	0.51	0.70	−0.87	−0.58	−0.66	0.72	−0.32	0.09
4	−0.50	−0.45	0.27	1.00	0.50	−0.02	0.20	−0.43	0.69	0.29	0.01	−0.09	0.00	0.22	−0.08	0.08
5	−0.47	−0.61	0.63	0.50	1.00	−0.32	0.50	−0.42	0.70	0.93	−0.47	−0.50	−0.36	0.79	−0.27	0.12
6	0.62	0.66	−0.79	−0.02	−0.32	1.00	−0.42	0.53	−0.35	−0.40	0.91	0.21	0.69	−0.40	0.09	−0.06
7	−0.47	−0.43	0.56	0.20	0.50	−0.42	1.00	−0.08	0.38	0.52	−0.48	−0.32	0.00	0.44	0.35	0.74
8	0.77	0.82	−0.69	−0.43	−0.42	0.53	−0.08	1.00	−0.54	−0.45	0.53	0.19	0.74	−0.41	0.48	0.36
9	−0.63	−0.65	0.51	0.69	0.70	−0.35	0.38	−0.54	1.00	0.62	−0.44	0.00	−0.34	0.32	0.01	0.12
10	−0.52	−0.66	0.70	0.29	0.93	−0.40	0.52	−0.45	0.62	1.00	−0.57	−0.50	−0.44	0.82	−0.27	0.10
11	0.60	0.71	−0.87	0.01	−0.47	0.91	−0.48	0.53	−0.44	−0.57	1.00	0.32	0.76	−0.54	0.16	−0.07
12	0.11	0.32	−0.58	−0.09	−0.50	0.21	−0.32	0.19	0.00	−0.50	0.32	1.00	0.28	−0.73	0.62	0.04
13	0.42	0.61	−0.66	0.00	−0.36	0.69	0.00	0.74	−0.34	−0.44	0.76	0.28	1.00	−0.43	0.59	0.45
14	−0.42	−0.58	0.72	0.22	0.79	−0.40	0.44	−0.41	0.32	0.82	−0.54	−0.73	−0.43	1.00	−0.44	0.00
15	0.04	0.27	−0.32	−0.08	−0.27	0.09	0.35	0.48	0.01	−0.27	0.16	0.62	0.59	−0.44	1.00	0.68
16	−0.14	−0.04	0.09	0.08	0.12	−0.06	0.74	0.36	0.12	0.10	−0.07	0.04	0.45	0.00	0.68	1.00

follows: The more number of trees to be obtained, it reported the higher the accuracy of the algorithm, but it has also been found that too many trees sometimes causes the over-fitting problem. In this work, we have considered the tree range from 1 to 100 with an interval of 5, and other parameters are default to sleep staging. Similarly, the second layer also employed one classification model (XGBoosting), which carried the input as the combinations of the prediction results from the base-layer classification model, as the new training datasets. Similarly, the combinations of the first layer's test results being treated on the new testing dataset for the second layer of the proposed models. In this work, we use the cross-validation and grid-search method to obtaining the optimal parameters. The maximum number of features (max_features) is considered as 50, the learning rate is 0.001, the maximum depth(max-depth) of the decision tree is 7, the maximum number of the iterations (n_estimators) for the weak learner is 5000, the minimum number of leaves (min_samples_leaf) is 90, the minimum numbers of samples required for internal node redistribution (min_samples_split) is 1000, and the other parameters are default. Finally, the *meta*-classifier layer provides the final sleep-staging prediction results with the new testing dataset. In summary, the proposed model obtained a two-layer ensemble learning stacking algorithm to improve the classification performance of the sleep staging.

2.6. Performance evaluation metrics

In this proposed study, we evaluated the performance of the model using the cross-validation strategies by considering performance metrics such as overall accuracy [71], sensitivity [72], precision [73], F1Score [74], and Cohen's Kappa coefficient [75]. It has been reported that evaluating the model's performance using a single dataset may sometimes lead to biased results due to the same recording procedures and conditions. To better assess the model's efficiency, our proposed model obtained four public datasets such as S-EDF, SE-EDF, Dreams subjects, and ISRUC-Sleep datasets, also obtaining both the sleep scoring rules such as R&K rules and AASM rules. To better analyze sleep behavior changes, we considered both the EEG montage recordings(monopolar and bipolar). R&K sleep scoring rules were used on S-EDF and SE-EDF datasets. Similarly, AASM scoring rules were obtained on the Dream and ISRUC-Sleep dataset for sleep staging Eqs. (8)–(13) shows the evaluation methods for these performance evaluation metrics.

$$Accuracy(ACC) = \frac{TP + TN}{TP + FN + TN + FP}\% \tag{8}$$

$$Sensitivity(SEN) = \frac{TP}{TP + FN}\% \tag{9}$$

$$Specificity(SPC) = \frac{TN}{TN + FP}\% \tag{10}$$

$$Precision(PRE) = \frac{TP}{TP + FP}\% \tag{11}$$

where:

TP = True positives
FN = False negatives
TN = True negatives
FP = False positives

$$F1Score(FSc) = \frac{2{*}Recall{*}Precision}{Recall + Precision} \tag{12}$$

The Kappa coefficient provides a means of measuring the best performance of the classification techniques. The Kappa score can be divided into six levels: excellent agreement (0.81–1), substantial agreement (0.61–0.80), moderate agreement (0.41–0.60), fair agreement (0.21–0.4), slight agreement (0–0.20) and poor agreement (<0) [76]

$$KappaCoefficient(KappaSc) = \frac{2{*}(TN{*}TP - FP{*}FN)}{(TN+FN){*}(FN+TP)+(FP+TP){*}(TN+FP)} \tag{13}$$

The authors use k cross-validation techniques to measure the performance of the proposed classification method. This metric estimates the quality of a classification method by dividing the number of correctly classified results by the total number of cases. This process is done five times, and the k-fold results are averaged and reported as the system performance.

2.7. Testing schemes

2.7.1. Subject independent test

In this study, we obtained k-fold cross-validation techniques used to evaluate the proposed model's performance. For the k-subjects, the model is repeated for the k-times, where one fold of the dataset is used to test while the other k-1 folds are used for training the proposed model.

2.7.2. Multi-class sleep staging evaluation

This research work reported the multi-class sleep staging results for all the four public datasets based on the R&K and AASM sleep scoring rules. According to the R&K rules, the sleep staging results for six sleep (CT-6) classification task (W, S1, S2, S3, S4, and REM), for five sleep (CT-5) classification task (W, S1, S2, S3 + S4, and REM), for four sleep (CT-4) classification task (W, S1 + S2, S3 + S4, and REM), for three sleep (CT-3)

Table 10
Pearson correlation matrix between first 18 features.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	−0.77	0.30	0.27	0.47	−0.54	−0.53	0.76	0.24	−0.43	−0.47	−0.49	0.26	0.30	0.14	0.25	−0.06	−0.06	−0.04
2	−0.77	1.00	−0.80	−0.71	−0.83	0.55	0.71	−0.55	−0.65	0.63	0.59	0.65	−0.63	−0.62	0.26	0.24	0.19	0.19	0.18
3	0.30	−0.80	1.00	0.89	0.71	−0.35	−0.60	0.21	0.53	−0.53	−0.42	−0.48	0.82	0.52	−0.38	−0.57	−0.15	−0.15	−0.14
4	0.27	−0.71	0.89	1.00	0.61	−0.25	−0.43	0.21	0.45	−0.35	−0.29	−0.40	0.75	0.49	−0.28	−0.37	−0.07	−0.07	−0.06
5	0.47	−0.83	0.71	0.61	1.00	−0.44	−0.75	0.36	0.84	−0.64	−0.62	−0.53	0.57	0.78	−0.53	−0.41	−0.29	−0.29	−0.27
6	−0.54	0.55	−0.35	−0.25	−0.44	1.00	0.62	−0.38	−0.29	0.74	0.58	0.37	−0.27	−0.29	0.02	−0.01	0.11	0.11	0.09
7	−0.53	0.71	−0.60	−0.43	−0.75	0.62	1.00	−0.48	−0.46	0.74	0.59	0.32	−0.54	−0.51	0.24	0.34	0.22	0.22	0.18
8	0.76	−0.55	0.21	0.21	0.36	−0.38	−0.48	1.00	0.08	−0.26	−0.27	0.00	0.46	0.45	0.22	0.25	0.38	0.38	0.41
9	0.24	−0.65	0.53	0.45	0.84	−0.29	−0.46	0.08	1.00	−0.48	−0.49	−0.52	0.32	0.74	−0.59	−0.40	−0.32	−0.32	−0.32
10	−0.43	0.63	−0.53	−0.35	−0.64	0.74	0.74	−0.26	−0.48	1.00	0.84	0.52	−0.40	−0.42	0.44	0.44	0.29	0.29	0.26
11	−0.47	0.59	−0.42	−0.29	−0.62	0.58	0.59	−0.27	−0.49	0.84	1.00	0.59	−0.28	−0.43	0.56	0.24	0.35	0.35	0.33
12	−0.49	0.65	−0.48	−0.40	−0.53	0.37	0.32	0.00	−0.52	0.52	0.59	1.00	−0.13	−0.19	0.34	0.19	0.67	0.67	0.67
13	0.26	−0.63	0.82	0.75	0.57	−0.27	−0.54	0.46	0.32	−0.40	−0.28	−0.13	1.00	0.64	−0.23	−0.45	0.22	0.22	0.25
14	0.30	−0.62	0.52	0.49	0.78	−0.29	−0.51	0.45	0.74	−0.42	−0.43	−0.19	0.64	1.00	−0.48	−0.33	0.12	0.12	0.15
15	0.14	0.26	−0.38	−0.28	−0.53	0.02	0.24	0.22	−0.59	0.44	0.56	0.34	−0.23	−0.48	1.00	0.68	0.31	0.31	0.31
16	0.25	0.24	−0.57	−0.37	−0.41	−0.01	0.34	0.25	−0.40	0.44	0.24	0.19	−0.45	−0.33	0.68	1.00	0.19	0.19	0.19
17	−0.06	0.19	−0.15	−0.07	−0.29	0.11	0.22	0.38	−0.32	0.29	0.35	0.67	0.22	0.12	0.31	0.19	1.00	1.00	0.99
18	−0.06	0.19	−0.15	−0.07	−0.29	0.11	0.22	0.38	−0.32	0.29	0.35	0.67	0.22	0.12	0.31	0.19	1.00	1.00	0.99

classification task, for two sleep (CT-2) classification task (W) and Sleep (NREM + REM)). These sleep class classification tasks have been mostly obtained for the evaluation of the sleep staging. Similarly, for the datasets using AASM sleep scoring rules, the sleep staging evaluation was performed for the five-class(W, N1, N2, N3, and REM), four-class (W, N1 + N2, N3, and REM), three-class (W, N1 + N2 + N3, and REM), two-class (W, and Sleep (NREM + REM)) cases. The entire execution was conducted using MATLAB R2017b software running on a personal laptop with an Intel Core™ i3-4005U CPU 1.70 GHz, two cores, four logical processors, 4 GB RAM, and Windows 10 operating system.

3. Experimental results

This sleep staging study performs eight groups of experiments based on the R&K sleep scoring rules. The brief settings for all the expressions are shown in Table 11. Each experiment was implemented in two different ways: one set of experiments done, including the feature age, and the other experiments that excluded the feature age. Experiment-1 to Experiment-8 use the R&K sleep scoring rules, six-two (CT-6 to CT-2) sleep states classification task, and 5-fold cross-validation techniques.

We have considered similar conditions like selecting channels, dataset size, and test conditions similar to the existing studies on the obtained two public datasets, which alternatively used during the final result analysis of the proposed model with the existing studies. Experiment-1 and Experiment-3 perform on the S-EDF dataset using the Fpz-Cz and Pz-Oz channels of the EEG signals with 49 features, respectively. To analyze the impacts of the subjects' age on the sleep staging classification performance, we conducted Experiment-2 and Experiment-4 without age features on the same S-EDF dataset and channels, respectively. Similarly, Experiment-5 and Experiment-7 perform on the SE-EDF dataset with the R&K sleep scoring rules including age (49) features; on the other hand, Experiment-6 and Experiment-8 perform without age (48) features.

3.1. Sleep staging results in S-EDF database

For Experiment-1 and Experiment-3, first of all, we obtained the ReliefF weight feature selection algorithm for screening the suitable features for the classification model from 15, 170 samples with 49 features and forwarded into the ensemble learning stacking model for classifying the sleep stages. The reported confusion matrix and performance evaluation results for the Experiment-1 and Experiment-2 are described in Tables 12 and 13. It has been observed from Table 12 that the Fpz-Cz channel of the EEG signal with the inclusion of age (49) features. The overall accuracy, F1score and kappa score of the proposed model, are 91.10%, 85.42%, and 0.87, and respectively, similarly, the same model achieved overall accuracy (90.68%), F1score (84.86%), and kappa score (0.86) with the exclusion of the age (48) features. On the other hand, Experiment-3 and Experiment-4 were based on the S-EDF dataset using the Pz-Oz EEG channel based on (or) without age features. The other experimental conditions are the same as Experiment-1 and Experiment-2. The confusion matrix and the performance metrics results for the Pz-Oz channel of the EEG signal are presented in Tables 14 and 15.

It has been observed from Tables 14 and 15 that the overall accuracy performance decreases from 90.56% to 90.11%. F1score decreases from 84.53% to 83.91%, and the kappa score remain the same with concerns to inclusion and exclusion of the age features, respectively. The performance for six-two class classification problems using selected features, which are screened through ReliefF for the S-EDF database, is shown in Fig. 13.

3.2. Sleep staging results in SE-EDF database

Experiment-5, Experiment-6, Experiment-7, and Experiment-8 are performed in this section to analyze the feasibility and robustness of the

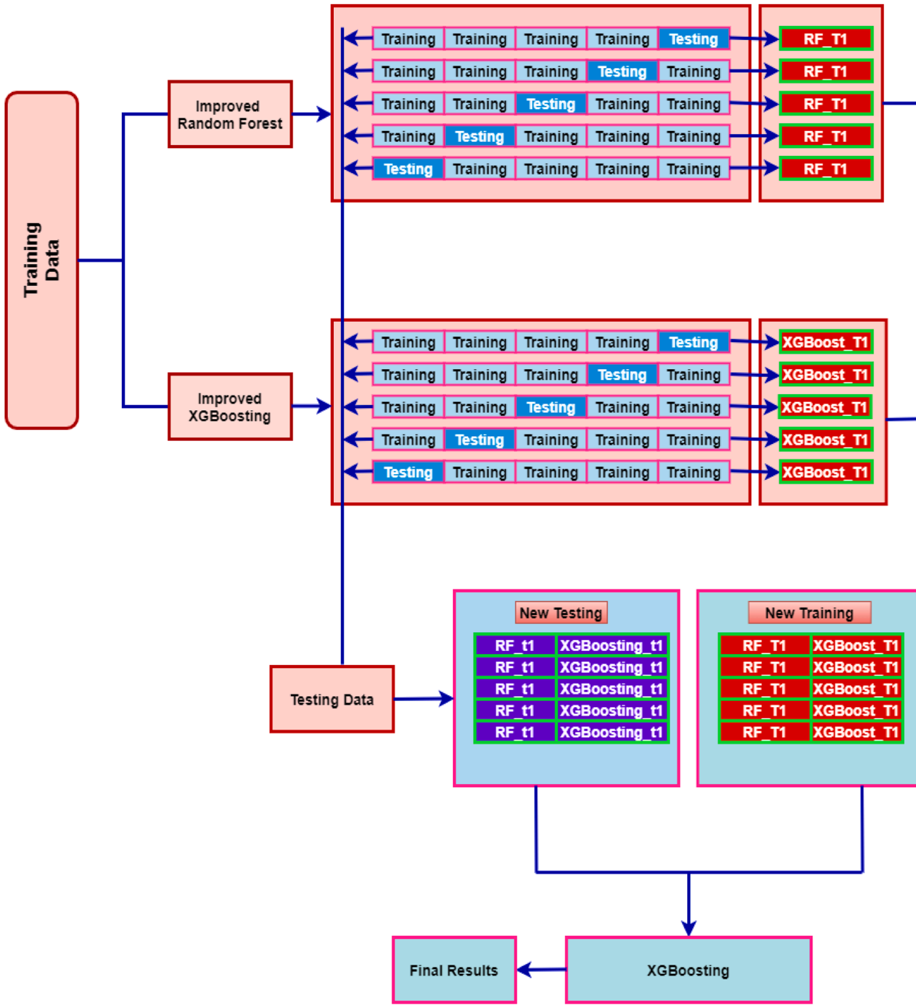


Fig. 12. Proposed Ensemble learning stacking model structure.

Table 11
Proposed experiments and their conditions.

Experiments	Datasets	Channels	No. of Features
Experiment-1	S-EDF	Fpz-Cz	49
Experiment-2	S-EDF	Fpz-Cz	48
Experiment-3	S-EDF	Pz-Oz	49
Experiment-4	S-EDF	Pz-OZ	48
Experiment-5	SE-EDF	Fpz-Cz	49
Experiment-6	SE-EDF	Fpz-Cz	48
Experiment-7	SE-EDF	Pz-Oz	49
Experiment-8	SE-EDF	Pz-Oz	48

proposed model like earlier experiments. The same parameters were obtained for training the model, sleep scoring rules, and cross-validation techniques. Experiment-5 and Experiment-6 were performed on the Fpz-Cz channel. Similarly, Experiment-7 and Experiment-8 performed on the Pz-Oz channel of the EEG signal with and without age features. The confusion matrix and the classification performance for the Experiment-5 and Experiment-6 input channel Fpz-Cz of EEG signal are shown in Tables 16 and 17. Similarly, for Experiment-7 and Experiment-8, the confusion matrix and sleep staging performance are presented in Tables 18 and 19 for the channel Pz-Oz with and without consideration of the age feature.

The reported performance metrics results for the Fpz-Cz channel, the ACC, F1Sc, and Kappa SC of the proposed model 81.32%, 82.17%, and

Table 12
The confusion matrix and classification performance of six sleep states classification on S-EDF dataset using EEG Fpz-Cz channel with the inclusion of age features under the R&K standards.

Automatic Scoring	Expert Scoring								
	W	S1	S2	S3	S4	REM	SEN (%)	PRE (%)	F1Score (%)
W	6671	37	25	8	10	55	95.65	98.01	96.81
S1	100	1095	51	3	1	54	86.22	83.97	85.08
S2	53	35	3273	101	25	134	92.51	90.38	91.43
S3	34	6	94	1269	168	1	90.32	80.72	85.25
S4	14	11	40	18	834	10	79.65	84.96	82.21
REM	102	86	55	6	9	651	71.93	71.61	71.76
Overall							ACC:91.10%	F1Sc:85.42%	Kappa Sc:0.87

Table 13
The confusion matrix and classification performance of six sleep states classification on S-EDF dataset using EEG Fpz-Cz channel with the exclusion of age features under the R&K standards.

		Expert Scoring								
		W	S1	S2	S3	S4	REM	SEN (%)	PRE (%)	F1Score (%)
Automatic Scoring	W	6651	57	25	8	10	55	95.72	97.72	96.71
	S1	109	1082	51	3	5	54	84.13	82.97	83.55
	S2	33	35	3253	131	35	134	93.90	89.83	91.82
	S3	34	10	69	1289	161	8	87.92	82.05	84.88
	S4	19	21	31	18	803	25	76.62	87.56	81.73
	REM	102	81	35	17	34	651	70.22	70.76	70.49
	Overall							ACC:90.68%	F1Sc:84.86%	Kappa Sc:0.86

Table 14
The confusion matrix and classification performance of six sleep states classification on S-EDF dataset using EEG Pz-Oz channel with the inclusion of age features under the R&K standards.

		Expert scoring								
		W	S1	S2	S3	S4	REM	SEN (%)	PRE (%)	F1Score (%)
Automatic Scoring	W	6714	95	49	11	45	12	98.17	96.39	97.27
	S1	48	1412	47	15	6	70	76.57	88.36	82.04
	S2	27	110	3033	55	127	69	91.71	88.65	90.15
	S3	16	12	115	1051	85	10	90.68	81.53	85.86
	S4	10	15	48	10	515	29	63.26	82.13	71.47
	REM	24	200	15	17	36	986	83.84	77.15	80.36
	Overall							ACC:90.56%	F1Sc:84.53%	Kappa Sc:0.86

Table 15
The confusion matrix and classification performance of six sleep states classification on S-EDF dataset using EEG Pz-Oz channel with the exclusion of age features under the R&K standards.

Automatic Scoring		Expert Scoring								
		W	S1	S2	S3	S4	REM	SEN (%)	PRE (%)	F1Score (%)
	W	6690	92	53	24	43	14	98.02	96.73	97.37
	S1	41	1409	47	22	11	68	75.75	88.17	81.49
	S2	37	110	3013	55	132	74	91.60	88.07	89.80
	S3	16	33	114	1031	85	10	88.19	79.98	83.88
	S4	17	21	42	13	512	22	62.51	81.65	70.81
	REM	24	195	20	24	36	979	83.89	76.60	80.08
	Overall							ACC:90.11%	F1Sc:83.91%	Kappa Sc:0.86

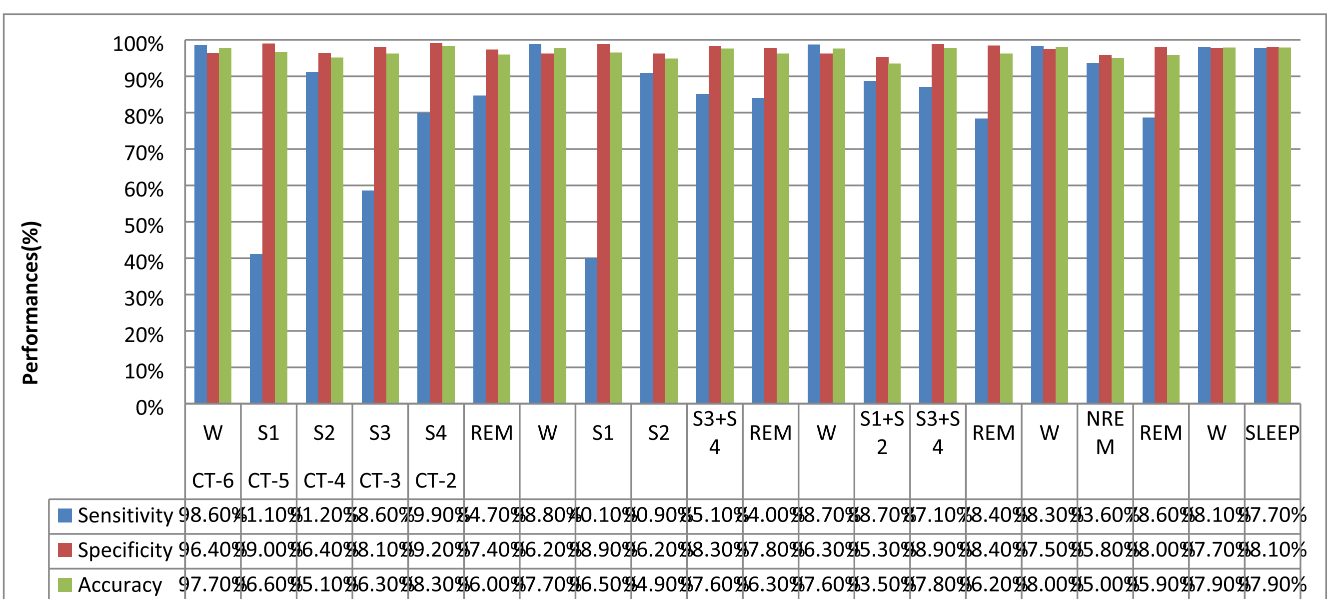


Fig. 13. The average performance of the proposed system for CT-6, CT-5, CT-4, CT-3, and CT-2 classification on S-EDF database obtained selected features using ReliefF feature screening algorithm.

Table 16
The confusion matrix and classification performance of six sleep states classification on SE-EDF dataset using EEG Fpz-Cz channel with the inclusion of age features under the R&K standards.

		Expert Scoring								
		W	S1	S2	S3	S4	REM	SEN (%)	PRE (%)	F1Score (%)
Automatic Classification	W	5330	176	214	18	16	323	77.40	87.70	82.23
	S1	581	6042	521	10	25	1025	90.59	73.64	81.24
	S2	451	132	8040	720	45	900	82.70	78.14	80.36
	S3	54	23	20	4810	413	50	83.33	89.57	86.34
	S4	39	18	65	201	2996	25	85.52	89.59	87.51
	REM	431	278	861	13	8	6126	72.50	78.38	75.33
	Overall							ACC:81.32%	F1Sc:82.17%	Kappa Sc:0.77

Table 17
The confusion matrix and classification performance of six sleep states classification on SE-EDF dataset using EEG Fpz-Cz channel with the exclusion of age features under the R&K standards.

		Expert Scoring								
		W	S1	S2	S3	S4	REM	SEN (%)	PRE (%)	F1Score (%)
Automatic Classification	W	5323	170	216	25	22	321	77.11	87.59	82.02
	S1	591	6022	522	20	28	1021	90.29	73.40	80.97
	S2	445	138	8025	735	53	892	82.63	78.00	80.25
	S3	64	33	22	4790	413	48	82.67	89.19	85.81
	S4	39	28	65	201	2976	35	84.88	88.95	86.87
	REM	441	278	861	23	14	6100	72.47	79.04	75.61
	Overall							ACC:80.51%	F1Sc:81.22 %	Kappa Sc:0.76

Table 18
The confusion matrix and classification performance of six sleep states classification on SE-EDF dataset using EEG Pz-Oz channel with the inclusion of age features under the R&K standards.

		Expert Scoring								
		W	S1	S2	S3	S4	REM	SEN (%)	PRE (%)	F1Score (%)
Automatic Scoring	W	5423	62	221	33	27	311	79.91	89.23	84.31
	S1	569	6012	542	22	28	1031	91.68	73.28	81.45
	S2	445	138	8025	735	53	892	82.87	78	80.36
	S3	69	39	29	3785	406	42	77.26	86.61	81.67
	S4	39	28	65	101	2975	135	83.47	88.99	86.14
	REM	241	278	801	223	75	6200	72.00	79.30	75.47
	Overall							ACC:80.84%	F1Sc:81.57%	Kappa Sc:0.76

Table 19
The confusion matrix and classification performance of six sleep states classification on SE-EDF dataset using EEG Pz-Oz channel with the exclusion of age features under the R&K standards.

		Expert Scoring								
		W	S1	S2	S3	S4	REM	SEN (%)	PRE (%)	F1Score (%)
Automatic Scoring	W	5403	66	231	38	32	307	79.72	88.90	84.06
	S1	549	6007	536	27	54	1031	91.12	73.22	81.20
	S2	431	145	8005	745	60	902	82.29	77.80	79.98
	S3	79	39	49	3745	411	47	77.61	85.69	81.45
	S4	64	53	65	51	2975	135	82.47	88.99	85.61
	REM	251	282	841	219	75	6150	71.74	78.66	75.04
	Overall							ACC:80.51%	F1Sc:81.22%	Kappa Sc:0.76

0.77, similarly for the channel Pz-Oz, the achieved results for the ACC (80.84%).F1Sc (81.57%), and Kappa Sc (0.76) using 49 features based on the SE-EDF dataset. The proposed model was reported with the same dataset without age features for both Fpz-Cz and Pz-Oz channels. The reported results indicate that the accuracy, F1Score, and Kappa score decreased from 80.84% to 80.51%, 81.57% to 81.22%, and 0.76 to 0.75, respectively. Finally, Table 20 presents the classification performance results for the six-class (CT-6), five-class (CT-5), four-class (CT-4), three-class (CT-3), and two-class (CT-2) based on the S-EDF and SE-EDF dataset. The reported performance for six to two sleep classes with selected features using the ReliefF feature selection algorithm using the SE-EDF database is shown in Fig. 14. The reported overall accuracy using a different number of features with S-EDF and SE-EDF databases is shown in Fig. 15.

4. Discussion

This proposed research work proposes a two-layer ensemble learning stacking model for multi-class sleep staging classification using EEG signals under the R&K rules. The proposed two-layer ensemble learning stacking model integrates the advantages of RF and XGBoost. The RF helps reduce the variance, while XGBoost reduces the bias of the proposed model. It has been found that the proposed model reported a promising performance incomparable to the existing contribution. The

Table 20
Overall accuracy performance and kappa score of the proposed model using S-EDF and SE-EDF datasets under the R&k sleep scoring rules.

Performance metrics	Dataset	Channel	No. of Feature	CT-6 (%)	CT-5 (%)	CT-4 (%)	CT-3 (%)	CT-2 (%)
Overall Accuracy	S-EDF	Fpz-Cz	49	91.10%	92.79%	93.66%	94.91%	97.88%
			48	90.68%	91.05%	91.94%	93.49%	96.77%
		Pz-Oz	49	90.56%	91.15%	92.37%	93.87%	96.63%
			48	90.11%	90.92%	91.78%	92.52%	94.72%
	SE-EDF	Fpz-Cz	49	81.32%	83.20%	85.29%	88.93%	95.48%
			48	81.06%	82.37%	84.10%	87.62%	94.39%
		Pz-Oz	49	80.84%	81.79%	84.02%	86.91%	92.71%
			48	80.51%	81.08%	83.96%	85.92%	92.05%
Cohen's kappa coefficient	S-EDF	Fpz-Cz	49	0.87	0.88	0.90	0.91	0.96
			48	0.86	0.87	0.89	0.90	0.94
		Pz-Oz	49	0.86	0.88	0.89	0.91	0.95
			48	0.86	0.87	0.88	0.90	0.93
	SE-EDF	Fpz-Cz	49	0.77	0.80	0.81	0.82	0.85
			48	0.76	0.77	0.89	0.81	0.83
		Pz-Oz	49	0.76	0.78	0.79	0.81	0.84
			48	0.76	0.77	0.79	0.80	0.82

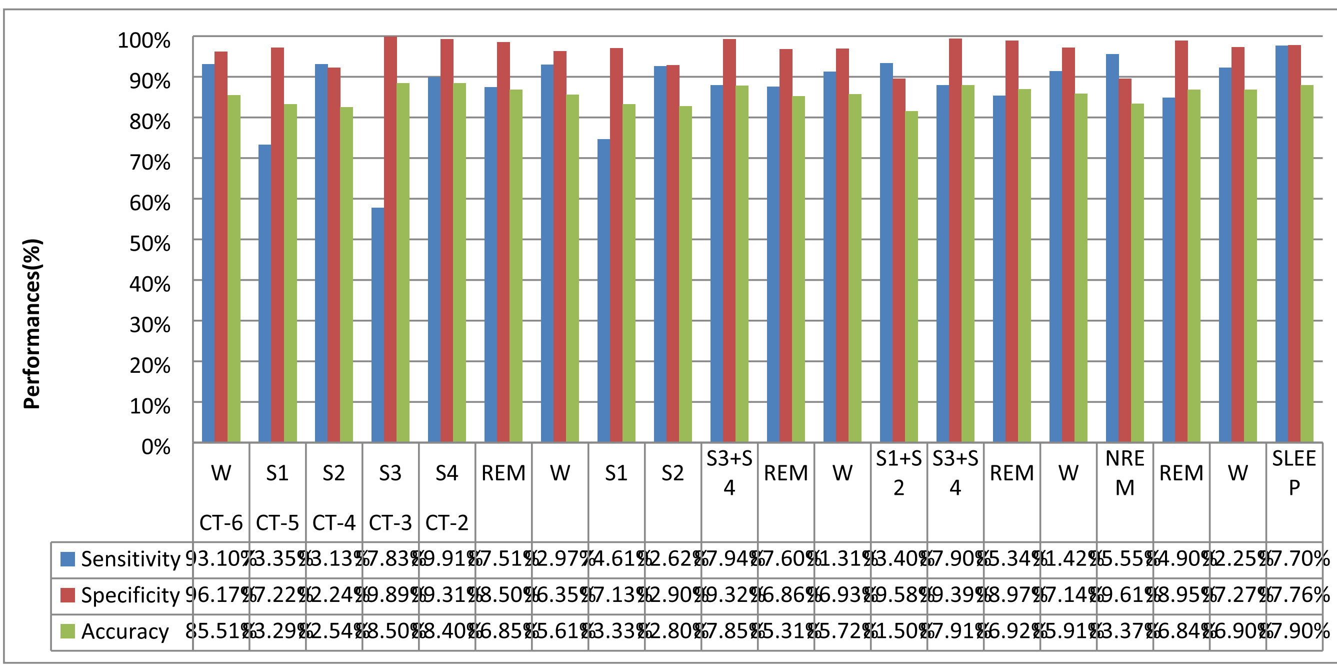


Fig. 14. Average performance of the proposed system for CT-6, CT-5, CT-4, CT-3, and CT-2 classification on SE-EDF database obtained selected features using ReliefF feature screening algorithm.

proposed ensemble learning stacking model improves the sleep staging performance concerning multi-class sleep staging, one of the common challenges for automated sleep staging systems. The proposed model particularly improves sleep staging discrimination, which is a common challenge in automated sleep staging. It has been observed that the recognition rate of the S1 (N1) stage is one of the major challenging tasks. Mainly there are two major reasons with subject to proper recognition of the N1 stage. Generally, the N1 sleep stage is the transaction stage between the wake and N2 stage. Sometimes the automated sleep staging system mispredicts it as wake (or) N2. It has also been noticed that there is a strong similarity between the N1 and REM stage. During the REM sleep stage, the gamma (γ) wave patterns are visible in the wake stages. It has been seen that the S1(N1) stage is often mislabelled as the REM (or) wake stage during the visual inspection of the sleep behavior of the subjects by the sleep experts. This second most reason is that human sleep is a combination of the different sleep stages with an unbalanced distribution of the sleep epochs into the different sleep stages. It has been found that the S1(N1) stage epochs are very less than in comparison to the other sleep stages, which seriously affects regards the proper recognition of the N1 sleep stage.

In this work, we obtained the class balance strategy, which indirectly improves the S1 (N1) sleep stage's recognition rate compared to earlier contributions. Moreover, in this proposed model, the other most important advantage is considering the age of the subject as one of the features, which plays an important role during the automated sleep staging system. In most of the earlier contributions, the authors have not addressed this same information in their feature extraction process. It has been noticed that the performance result from Table 12, Table 14, Tables 16, and 18 has proved that there should be a positive impact on the performance of the sleep staging system with the inclusion of the age factor of the subject as one of the feature. Furthermore, the performance of the proposed model in terms to accuracy, F1Score, and Kappascore. Additionally, the EEG signals' channels greatly help analyze the subjects' sleep behavior and simplify the sleep scoring system.

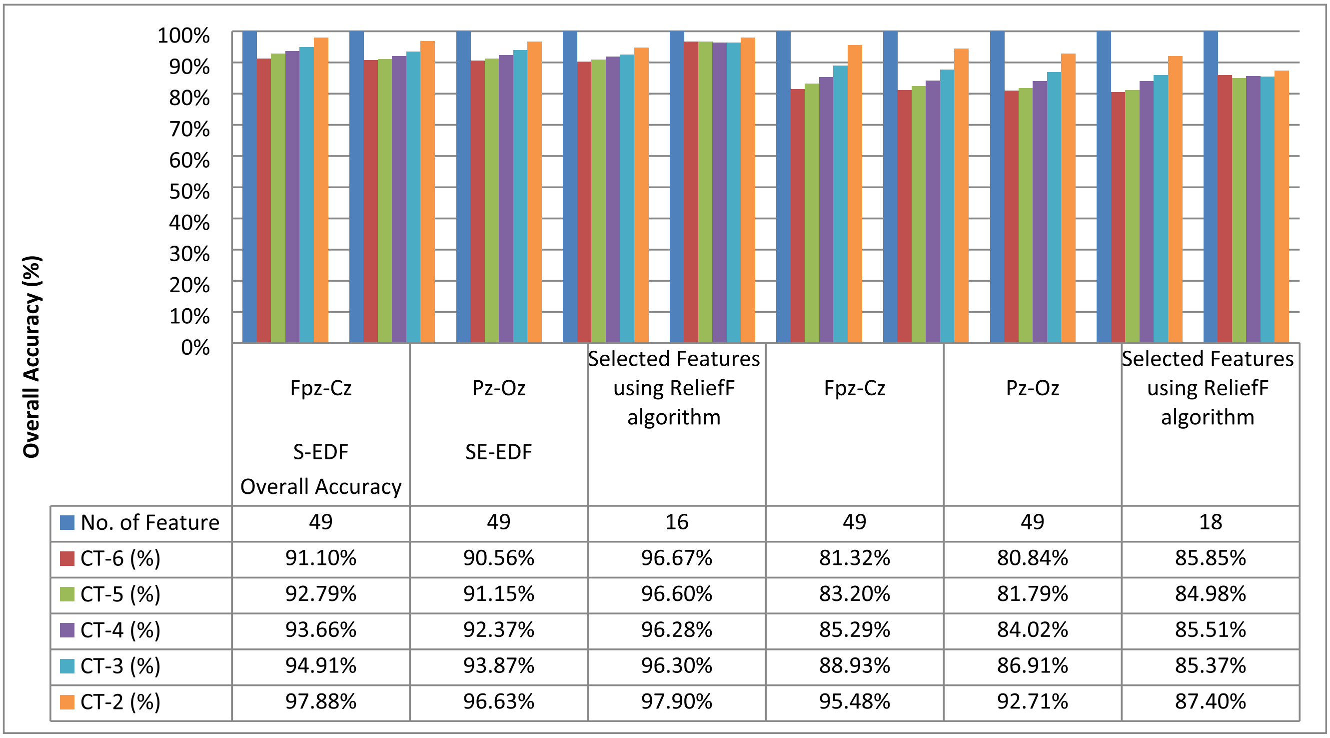


Fig. 15. Overall accuracy performance of the proposed system for CT-6, CT-5, CT-4, CT-3 and CT-2 classification on SE-EDF database using different numbers of features.

4.1. Comparison of the results of the proposed model with other existing contribution to the sleep-staging system

The performance of the proposed model compared with the existing sleep staging contribution into the different aspects. In this research work, we have analyzed the proposed model's results for the individual dataset. Table 21 presents the performance comparison results of the proposed model in terms of accuracy, F1score, Kappa score, and sensitivity of the S1 stage with state-of-the-art works on the S-EDF dataset six-sleep states classification under the R&K sleep scoring rules. Finally, Table 22 presents the performance comparison results of the various earlier sleep staging studies under the R&K sleep scoring rules.

It has been observed from Table 21 that the proposed methodology in this paper has a certain improvement incomparable to the existing state-of-the-art works regarding five sleep states classification (CT-6) based on the S-EDF dataset under the R&K sleep standards. It has been found that the highest classification accuracy achieved with the Fpz-Oz channel of EEG signals from the proposed model, which is significantly improved accuracy incomparable to the earlier published contribution. Additionally, the proposed model reported a better recognition rate for the S1 sleep stage, which is always one of the major challenges during sleep staging. The reported results for the S1 sleep stage using Pz-Oz is 76.57%

Table 21
Performance metrics results for five-sleep states classification compared with published contributions on the S-EDF dataset.

Authors	Epoch Number	Channel	Methods	Accuracy	F1Score	Kappa Score	Sensitivity S1 stage
Hassan A.R. et al., 2015 Ref. [17]	15,188	Pz-Oz	Spectral features and AdaBoost	82%	81.6%	0.71	22.85%
Hassan A.R. et al., 2016 Ref . [3]	15,188	Pz-Oz	Statistical features and EMD	90.1%	89.7%	0.84	39.74%
Hassan A.R. et al., 2017 Ref. [77]	15,188	Pz-Oz	CEEMDAN and Bagging	90.7%	90.5%	0.85	47.02%
Hassan A.R. et al., 2016 Ref . [17]	15,188	Pz-Oz	TQWT and RF	91.5%	90.8%	0.86	37.42%
Hassan A.R. et al., 2017 Ref [78]	15,188	Pz-Oz	TQWT and AdaBoost	91.4%	91%	0.86	39.74%
Hassan A.R. et al., 2017 Ref [79]	15,188	Pz-Oz	RUSBoost	83.5%		0.84	42.05%
Hassan A.R. et al., 2017 Ref [80]	15,188	Pz-Oz	Statistical features and TQWT	90.8%	90.4%	0.85	38.74%
Jiang et al. 2019 Ref [14]	15,160	Pz-Oz	RF and HMM	92%	91.4%	0.87	–
		Fpz-Cz	RF and HMM	92.7%	92.2%	0.88	–
Huaming Shen et al. 2020 Ref [82]	10,4643	Pz-Oz	Improved Model based Essence Features(IMBEFs) + Dual State Space Models(DSSMs)	92.04%	–	0.82	–
Proposed Study	15,139	Pz-Oz	Proposed Method(with age feature)	91.15%	89.99%	0.86	76.57%
			Proposed Method(without age feature)	90.92%	89.01%	0.86	75.75%
		Fpz-Cz	Proposed Method(with age feature)	92.79%	91.77%	0.88	86.22%
			Proposed Method(with age feature)	91.05%	90.32%	0.87	84.13%

Table 22
The performance for six-two sleep classes classifications with the various published contribution works on the S-EDF dataset under R&K sleep scoring rules.

Authors	Epoch Number	Cross Validation	CT-6	CT-5	CT-4	CT-3	CT-2
Zhu et al., 2014 Ref [11]	14,963	10 Fold	87.5%	88.9%	89.3%	92.6%	97.9%
Hassan et al., 2016 Ref [17]	15,188	0.5/0.5	90.38%	91.50%	92.11%	94.8%	97.5%
Hassan et al., 2016 Ref [3]	15,188	05/0.5	88.62%	90.11%	91.2%	93.5%	97.73%
Hassan et al., 2017 Ref [79]	15,188	0.5/0.5	88.07%	83.49%	92.66%	94.23%	98.15%
Sharma et al., 2017 Ref [18]	15,139	10-Fold	90.03%	91.13%	92.29%	94.66%	98.02%
Supratak et al., 2017 Ref [83]	41,950	20-Fold	–	79.8%	–	–	–
Sharma et al., 2018 Ref [84]	85,900	10-Fold	91.5%	91.7%	92.1%	93.9%	98.3%
Rahman et al., 2018 Ref [85]	15,188	0.5/0.5	90.26%	91.02%	92.89%	94.1%	98.24%
Abdulla et al., 2019 Ref [86]	23,806	–	93%	–	–	–	–
Ghimatgar et al., 2019 Ref [81]	15,188	0.5/0.5	89.91%	91.11%	92.19%	94.65%	98.19%
Ghimatgar et al., 2019 Ref [81]	40,100	0.5/0.5	79.13%	81.86%	83.71%	88.37%	95.98%
Michielli et al., 2019 Ref [87]	10,280	10 Fold	–	86.7%	–	–	–
Proposed Study (With Age)	15,139	5 Fold	91.10%	91.70%	92.66%	94.91%	97.88%
Proposed Study (Without Age)			90.68%	91.05%	91.94%	93.49%	96.77%

and 75%0.75%. Similarly, for the channel Fpz-Cz, the S1 stage achieved 86.22% and 84.13% with the inclusion and exclusion of the age feature, respectively. Table 22 provides the performance comparison results of the proposed model with the various related published contributions. It has been noticed that the proposed model performed well for the CT-6, CT-5, CT-4, and CT-3 sleep stages classification. All the eight experiments of this research work performed with different conditions report the proposed model's effectiveness regarding the automated sleep staging system. However, several advantages of this proposed model could be realized with the automated sleep staging under the R&K and AASM sleep scoring rules. It has also been noticed some of the limitations of the currently proposed methodology need to handle further in our research work. Firstly, the proposed study has not analyzed EEG signals' depth and not extracts the more distinctive features from the signals.

This study only handled the EEG signals decomposed into nine signal sub-bands and used handcrafted features in the earlier literature. The only new feature is considered the subject's age for proper discrimination of the S1(N1) sleep stage. Further, we will also focus on the proper recognition through deep analysis of the S1 (N1) and REM sleep stage characteristics, which ultimately improve the accuracy rate of the S1 (N1) sleep stage under both AASM and R&K, sleep scoring rules. Secondly, the imbalance of the sleep epochs' distribution to the different sleep stages, which sometimes gives the biased sleep staging performance, we will further resolve this problem by introducing the data augmentation techniques. So further research can stress the issues mentioned above to improve the model's sleep staging classification performance.

5. Conclusions

This proposed study considers developing an automated sleep staging classification system with dual-channel EEG signals. To improve the sleep staging performance, the proposed method presents a two-layer ensemble learning stacking model. The proposed model was evaluated on the two public datasets such as S-EDF and SE-EDF. The most important part of this research work is the inclusion of the subject's age as one of the features, ultimately providing better results for all the obtained datasets. The experimental results of the proposed model signify that the proposed methodology using performed using S-EDF dataset, the overall accuracy and Cohen's kappa coefficient score obtained by the proposed model is (91.10%, 0.87) and (90.68%, 0.86) with inclusion and age exclusion feature using the Fpz-Cz channel, respectively. Similarly, the Pz-Oz channel's performance is (90.56%, 0.86) with age feature and (90.11%, 0.86) without age feature. The performed results with the SE-EDF dataset using the Fpz-Cz channel are (81.32%, 0.77) and (81.06%, 0.76), using the Pz-Oz channel with the inclusion and exclusion of the age feature, respectively. Further, the proposed model was tested with the selected features obtained using the feature screening algorithm as the ReliefF algorithm. The model achieved an overall accuracy of 96.67% (CT-6), 96.60% (CT-5), 96.28% (CT-4),96.30% (CT-3) and 97.30% (CT-2) for with 16 selected features using S-EDF database. Similarly, the model reported an overall accuracy of 85.85%,84.98%,85.51%,85.37%, and 87.40% for ct-6 to ct-2 with 18 selected features using the SE-EDF database under R&K sleep scoring rules. The proposed work could be extended by integrating with an automated system based on a stacking ensemble approach based deep neural network model. The proposed model's reported results show that the proposed model provides higher sleep staging accuracy incomparable to the existing state-of-the-art works. The proposed methodology shows significant improvement in the automated sleep staging performance, which supports clinical practices for the various types of sleep-related diseases.

Authors' contributions

All the authors have equally contributed to this manuscript.

Funding

"This research received no external funding."

Ethics approval

'Not applicable' as publicly available datasets were analyzed in this study.

CRediT authorship contribution statement

Santosh Kumar Satapathy: Conceptualization, Methodology, Software, Writing - original draft, Investigation. **Akash Kumar Bhoi:** Data curation, Writing - original draft, Software, Visualization. **D. Loganathan:** Visualization, Investigation, Supervision. **Bidita Khandelwal:** Software, Validation, Investigation. **Paolo Barsocchi:** Writing - review & editing, Investigation, Validation, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L.A. Panossian, A.Y. Avidan, Review of sleep disorders, Med. Clin. N. Am. 93 (2009) 407–425.

[2] A. Smaldone, J.C. Honig, M.W. Byrne, Sleepless in America: inadequate sleep and relationships to health and well-being of our nation's children, Pediatrics 119 (2007) 29–37.

[3] A.R. Hassan, M.I. Hassan Bhuiyan, Automatic sleep scoring using statistical features in the EMD domain and ensemble methods, Bio Cybernet. Biomed. Eng. 36 (1) (2016) 248–255.
[4] Aboalayon Khald, Ocbagabir Helen, Faezipour, Miad. Efficient Sleep Stage Classification Based on EEG Signals.2014 IEEE Long Island Systems, Applications and Technology Conference, LISAT 2014.
[5] Obayya Marwa, abou-chadi, Fatma. Automatic classification of sleep stages using EEG records based on Fuzzy c-means (FCM) algorithm. National Radio Science Conference, NRSC, Proceedings;2014, p. 265-272.
[6] E. Alickovic, A. Subasi, Ensemble SVM method for automatic sleep stage classification, IEEE Trans. Instrum. Measurement 67 (6) (2018) 1258–12657.
[7] U.R. Abeyratne, V. Swarnkar, S.I. Rathnayake, C. Hukins, Sleep-stage and event dependency of brain asynchrony as manifested through surface EEG, in: In Proceedings of the 29th IEEE Annual International Conference of the Engineering in Medicine and Biology Society, 2007, pp. 709–712.
[8] A. Rechtschaffen, A. Kales, A Manual of Standardized Terminology, Techniques and Scoring Systems for Sleep Stages of Human Subjects, Public Health Service; Washington, DC, USA, UGP Office, 1968.
[9] C. Iber, S. Ancoli-Israel, A.L. Chesson, S.F. Quan, The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification, American Academy of Sleep Medicine, 2007.
[10] S.F. Liang, C.E. Kuo, Y.H. Hu, Y.H. Pan, Y.H. Wang, Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models, IEEE Trans. Instrum. Meas. 61 (6) (2012) 16491657.
[11] G. Zhu, Y. Li, P. Wen, Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal, IEEE J. Biomed. Health Inform. 18 (6) (2014) 18131821.
[12] C.S. Huang, C.L. Lin, L.W. Ko, S.Y. Liu, T.P. Su, C.T. Lin, Knowledge-based identification of sleep stages based on two for eheadelectroencephalogram channels, Front Neurosci. 4 (8) (2014) 263.
[13] P. Memar, F. Faradji, A novel multi-class EEG-based sleep stage classification system, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (1) (2018) 84–95.
[14] D. Jiang, Y.N. Lu, Y. Ma, Y. Wang, Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement, Expert Syst. Appl. 121 (2018) 188–203.
[15] Ronzhina Marina, Sleep scoring using artificial neural networks, Sleep Med. Rev. 16 (3) (2012) 251–263.
[16] Y.L. Hsu, Y.T. Yang, J.S. Wang, C.Y. Hsu, Automatic sleep stage recurrent neural classifier using energy features of EEG signals, Neurocomputing 104 (2013) p. 105_114.
[17] Hassan Ahnaf Rashik, Bhuiyan Mohammed, Imamul Hassan, A decision support system for automatic sleep staging from EEG signals using tunable Q-Factor wavelet transform and spectral features, J. Neurosci. Methods 271 (2016) 107–118.
[18] R. Sharma, R.B. Pachori, A. Upadhyay, Automatic sleep stages classification based on iterative filtering of electroencephalogram signals, Neural Comput. Applic 28 (2017) 2959–2978.
[19] Seifpour Saman, Niknazar Hamid, Mikaeili Mohammad, Nasrabadi Ali Motie, A new automatic sleep staging system based on statistical behaviour of local extrema using single channel EEG signal, Expert Syst. Appl. 104 (2018) 277–293.
[20] Moradi Parham, Rostami Mehrdad, Integration of graph clustering with ant colony optimization for feature selection, Knowledge Based Syst. 84 (2015) 144–161.
[21] Kira K, Rendell LA. The feature selection problem: Traditional methods and a new algorithm. in Proc. 10th Nat. Conf. Artif. Intell. ;1992, p.129_134.
[22] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell 27 (8) (2005) 1226–1238.
[23] X. Jin, T. Bo, H. He, M. Hong, Semi supervised feature selection based on relevance and redundancy criteria, IEEE Trans Neural Netw Learn Syst. 28 (2016) 1974–1984.
[24] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (1–3) (2002) 389–422.
[25] D. Ververidis, C. Kotropoulos, Fast sequential floating forward selection applied to emotional speech features estimated on DES and SUSAS data collections.In, Signal Process. Conf. (2006) 1–5.
[26] L. Yu, H. Liu, Feature selection for high-dimensional data: a fast correlation-based filter solution, in: In: Twentieth International Conference on Machine Learning, 2003, pp. 856–863.
[27] J.R. Quinlan, C4.5: Programs for Machine Learning, 1st ed., Morgan Kaufmann, San Mateo, CA, USA, 1992, pp. 313–320.
[28] Crivello A, Barsocchi P, Girolami M, Palumbo F. The Meaning of Sleep Quality: a Survey of Available Technologies. IEEE Access, 2019; p.1–1.
[29] Crivello A, Palumbo F, Barsocchi P, La Rosa D, Scarselli F, Bianchini M. Understanding Human Sleep Behaviour by Machine Learning. In: Klempous R, Nikodem J, Baranyi P. (eds) Cognitive Info communications, Theory and Applications. Topics in Intelligent Engineering and Informatics; 2019,vol 13.
[30] Tsinalis Orestis, M. Matthews Paul, Guo Yike, Automatic sleep stage scoring using time-frequency analysis and stacked sparse auto encoders, Ann. Biomed. Eng. 44 (5) (2016) 1587–1597.
[31] Zhang Junming, Wu Yan, Bai Jing, Chen, Fuqiang. Automatic sleep stage classification based on sparse deep belief net and combination of multiple classifiers. Trans. Inst. Meas. Control. ;2016, 38 (4), p.435–451.
[32] Kern Simon Johannes. Automatic Sleep Stage Classification Using Convolutional Neural Networks with Long Short-term Memory; 2017.
[33] A. Patanaik, J.L. Ong, J.J. Gooley, S. Ancoli-Israel, M.W.L. Chee, An end-to-end framework for real-time automatic sleep stage classification, Sleep 41 (5) (2018).
[34] Supratak Akara, Dong Hao, Wu Chao, Yike, Guo. DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG. IEEE Trans. Neural Syst. Rehabil. Eng.;2017, 25 (11),p.1998-2008.
[35] A.L. Alfeo, P. Barsocchi, M.G.C.A. Cimino, et al., Sleep behavior assessment via smartwatch and stigmergic receptive fields, Pers Ubiquit Comput 22 (2018) 227–243.
[36] Güneş K Polat, Ş. Yosunkaya, Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting, Expert Syst. Appl. 37 (2010) 7922–7928.
[37] Aboalayon K, Ocbagabir HT, Faezipour M. Efficient sleep stage classification based on EEG signals. Systems, Applications and Technology Conference (LISAT);2014, p.1-6.
[38] P. Barsocchi, M. Bianchini, A. Crivello, D. La Rosa, F. Palumbo, F. Scarselli, An unobtrusive sleep monitoring system for the human sleep behaviour understanding, in: 2016 7th IEEE International Conference on Cognitive Info communications (CogInfoCom), 2016, pp. 91–96.
[39] M. Diykh, Y. Li, Wen P.EEG sleep stages classification based on time do-main features and structural graph similarity, IEEE Trans. Neural Syst. Rehabil. Eng. 24 (11) (2016) 1159–1168.
[40] Gunnarsdottir KM, Gamaldo CE, Salas RM. E, Ewen JB, Allen RP, Sarma SV.A Novel Sleep Stage Scoring System: Combining Expert-Based Rules with a Decision Tree Classifier.40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018.
[41] N. Sriraam, T.K. PadmaShri, U. Maheshwari, Recognition of wake-sleep stage 1 multi-channel EEG patterns using spectral entropy features for drowsiness detection, Austr. Phys. Eng. Sci. Med. 39 (3) (2018) 797–806.
[42] T.L.T. Da Silveira, A.J. Kozakevicius, C.R. Rodrigues, Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain, Med. Biol. Eng. Comput. 55 (2) (2016) 343–352.
[43] Wutzl B, Leibnitz K, Rattay F.,Kronbichler, M., Murata, M. Genetic algorithms for feature selection when classifying severe chronic disorders of consciousness. PLOS ONE;2019,14(7).
[44] Braun ET, Kozakevicius ADJ, Da Silveira TL. T, Rodrigues CR, Baratto G. Sleep Stages Classification Using Spectral Based Statistical Moments as Features. Revista de Informática Teórica e Aplicada, 25(1).
[45] Pan Shing-Tai Kuo, Chih-En Zeng, Jian-Hong, Liang.; Sheng-Fu.A transition constrained discrete hidden markov model for automatic sleep staging. Biomed. Eng.;2012,11 (1), p.52.
[46] Ghimatgar Hojat, Kazemi Kamra, Helfroush, Mohamamd Sadegh, Aarabi Ardalan. An improved feature selection algorithm based on graph clustering and ant colony optimization. Knowledge Based Syst.;2018.
[47] B. Kemp, A.H. Zwinderman, B. Tuk, H.A.C. Kamphuisen, J.J.L. Oberye, Analysis of a sleep-dependent neuronal feedback loop: The slow-wave micro continuity of the EEG.IEEE Trans, Biomed. Eng. 47 (9) (2000) 11851194.
[48] M.S. Mourtazaev, B. Kemp, A.H. Zwinderman, H.A.C. Kamphuisen, Age and gender affect different characteristics of slow waves in the sleep EEG, Sleep 18 (7) (1995) 557–564.
[49] D. Stephanie, K. Myriam, P. Stenuit, M. Kerkhofs, E. Stanus, Cancelling ECG artifacts in EEG using a modified independent component analysis approach, EURASIP J. Adv. Signal Process 1 (2008), 747325.
[50] S. Khalighi, T. Sousa, J.M. Santos, U. Nunes, ISRUC-Sleep: a comprehensive public dataset for sleep researchers, Comput. Methods Programs Biomed. 124 (2016) 180–192.
[51] B. Koley, D. Dey, An ensemble system for automatic sleep stage classification using single channel EEG signal, Comput. Biol. Med. 42 (12) (2012) 1186_1195.
[52] T. Lajnef, S. Chaibi, P. Ruby, P.E. Aguera, J.B. Eichenlaub, M. Samet, A. Kachouri, K. Jerbi, Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines, J. Neurosci. Methods 250 (94105) (2015).
[53] Khalighi S, Sousa T, Pires G. and Nunes U. Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels.Expert Syst. Appl.; 2013, 40(17) p.7046_7059.
[54] K. Yasoda, R.S. Ponmagal, K.S. Bhuvaneshwari, Automatic detection and classification of EEG artifacts using fuzzy kernel SVM and wavelet ICA (WICA), Soft. Comput. 2020 (24) (2020) 16011–16019.
[55] W. Ahmad, A. Ahmad, C. Lu, A novel hybrid decision support system for thyroid disease forecasting, Soft. Comput. 22 (2018) 5377–5383.
[56] A. Gupta, R.K. Agrawal, B. Kaur, Performance enhancement of mental task classification using EEG signal: a study of multivariate feature selection methods, Soft Comput. 19 (2015) 2799–2812.
[57] Ghayab Al, Li, HR, Siuly, YS Epileptic seizures detection in EEGs blending frequency domain with information gain technique. Soft Comput;2019, 23, p.227–239.
[58] S. Al-Janabi, M. Mohammad, A. Al-Sultan, A new method for prediction of air pollution based on intelligent computation, Soft Comput. (2019).
[59] Souri A,Ghafour MY,Ahmed AM.A new machine learning-based healthcare monitoring model for student's condition diagnosis in Internet of Things environment. Soft Comput;020 24, p.17111–17121.
[60] V. Schetinin, J. Schult, Learning polynomial networks for classification of clinical electroencephalograms, Soft Comput 10 (2006) 397–403.
[61] K. Susmáková, A. Krakovská, Discrimination ability of individual measures used in sleep stages classification, Artif Intell Med. 44 (3) (2008) 261–277.
[62] A. Krakovská, K. Mezeiová, Automatic sleep scoring: a search for an optimal combination of measures, Artif Intell Med. 53 (1) (2011) 25–33.

[63] Lukas Zoubek, Sylvie Charbonnier, Suzanne, Lesecq, Alain Buguet, Florian Chapotot. Feature selection for sleep/wake stages classification using data driven methods. Biomedical Signal Processing and Control; 2007, 2 (3), p.171-179.

[64] S. Charbonnier, L. Zoubek, S. Lesecq, F. Chapotot, Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging, Comput. Biol. Med. 41 (6) (2011) 380–389.

[65] Y.L. Hsu, Y.T.C. Yang, J.S. Wang, Hsu CY Automatic sleep stage recurrent neural classifier using energy features of EEG signals, Neurocomputing 104 (2013) 105–114.

[66] Susmáková K, Krakovská A. Discrimination ability of individual measures used in sleep stages classification.Artif Intell Med. 2008,44(3):p.261-77.

[67] S. Al-Baddai, P. Marti-Puig, E. Gallego-Jutglà, A recognition–verification system for noisy faces based on empirical mode decomposition with Green's functions, Soft Comput. 24 (2020) 3809–3827.

[68] Zoubek Lukáš, Charbonnier Sylvie, Lesecq, Suzanne, Buguet Alain, Chapotot Florian. Feature selection for sleep/wake stages classification using data driven methods.Biomedical Signal Processing and Control.;2007, 2. p.171-179.

[69] Robnik-Šikonja M, Kononenko I.Theoretical and Empirical Analysis of ReliefF and RReliefF. Machine Learning; 2003, 53, 23–69.

[70] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference On Knowledge Discovery and Data Mining; 2016, p.785–794.

[71] T.H. Sanders, M. McCurry, M.A. Clements, Sleep stage classification with cross frequency coupling.36th Annu, Int. Conf. IEEE Eng. Med. Biol. (EMBC) (2014) 4579–4582.

[72] V. Bajaj, R. Pachori, Automatic classification of sleep stages based on the time-frequency image of EEG signals, Comput. Methods Programs Biomed. 112 (3) (2013) 320–328.

[73] Y.L. Hsu, Y.T. Yang, J.S. Wang, C.Y. Hsu, Automatic sleep stage recurrent neural classifier using energy features of EEG signals, Neuro Comput. 104 (2013) 105–114.

[74] I. Zibrandtsen, P. Kidmose, M. Otto, J. Ibsen, T.W. Kjaer, Case comparison of sleep features from ear-EEG and scalp-EEG, Sleep Sci. 9 (2) (2016) 69–72.

[75] R.B. Berry, R. Brooks, C.E. Gamaldo, S.M. Hardsim, R.M. Lloyd, C.L. Marcus, B. V. Vaughn, The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, Am. Acad. Sleep Med. (2014).

[76] J. Sim, C.C. Wright, The kappa statistic in reliability studies: use, interpretation, and sample size requirements, Phys. Ther. 85 (3) (2005) 257–268.

[77] Hassan AR, Bhuiyan MIH.Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. Biomed. Signal Process 2016. Control, 24, 1–10.

[78] A.R. Hassan, M.I.H. Bhuiyan, An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting, Neuro Comput. 219 (2016) 76–87.

[79] A.R. Hassan, M.I.H. Bhuiyan, Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting, Comput. Methods Programs Biomed. 140 (2017) 201–210.

[80] Hassan AR, Subasi, A.A decision support system for automated identification of sleep stages from single-channel EEG signals. Knowl.-Based Syst.;2017,128, p.115_124.

[81] H. Ghimatgar, K. Kazemi, M.S. Helfroush, A. Aarabi, An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov model, J. Neurosci. Methods 324 (2019) 180320–180336.

[82] H. Shen, F. Ran, M. Xu, A. Guez, A. Li, A. Guo, An automatic sleep stage classification algorithm using improved model based essence features, Sensors 20 (17) (2020) 4677.

[83] Supratak A, Dong H, Wu C, Guo Y. A model for automatic sleep stage scoring based on raw single-channel EEG.IEEE T. Neur. Sys. Reh. ; 2017, 25, p.998–2008.

[84] M. Sharma, D. Goyal, P.V. Achuth, U.R. Acharya, An accurate sleep stage classification system using a new class of optimally time-frequency localized three-band wavelet filter bank, Comput. Biol. Med. 98 (2018) 58–75.

[85] M.M. Rahman, M.I.H. Bhuiyan, A.R. Hassan, Sleep stage classification using single-channel EOG, Comput. Biol. Med. 10 (2018) 211–220.

[86] S. Abdulla, M. Diykh, R.L. Laft, K. Saleh, R.C. Deo, Sleep EEG signal analysis based on correlation graph similarity coupled with an ensemble extreme machine learning algorithm, Expert Syst. Appl. 138 (2019) 112790–112804.

[87] N. Michielli, U.R. Acharya, F. Molinari, Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals, Comput. Biol. Med. 106 (2019) 71–81.