



Prognosis of automated sleep staging based on two-layer ensemble learning stacking model using single-channel EEG signal

Santosh Kumar Satapathy¹ · D. Loganathan¹

Accepted: 25 February 2021 / Published online: 7 September 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Sleep is important part for human health and quality of life in the daily routine basis. However, numerous individuals face sleep problems due to rapid changes occurred in both social and professional lifestyles. These problems can lead to several neurological and physical disorder diseases, and therefore, decrease their overall life quality. Machine learning methods for automated sleep stage classification (ASSC) are a fundamental approach to evaluate and treat this public health challenge. The main objective of this study is to propose a high-effective and high-accuracy based multiple sleep staging classification model based on single-channel electroencephalogram (EEG) signals using machine learning (ML) model. The proposed automated sleep staging system followed four basic stages: signal preprocessing, feature extraction and screening, classification algorithms, and performance evaluation. In this research work, a novel method is applied for signal preprocessing, feature screening and classification models. In signal preprocessing we obtain the wiener filter techniques for removing the different types of artifacts from input sleep recordings. In feature extraction, we obtain a total of 28 features based on both time and frequency domain features and non-linear features. The relevant features are screened through ReliefF weight feature selection algorithm, and eliminating the redundant features using Pearson correlation coefficients. The important contribution of this research work is establishes two layers an ensembling learning stacking model for classifying the multiple sleep stages. Three different subgroups of ISRUC-Sleep (SG-I/SG-II/SG-III) subjects sleep recordings having different health condition obtained for our proposed experimental work. Comparing with the recent contributions on sleep staging performances using single-channel EEG signals, it has found that our proposed ensemble learning stacking model was reported excellent in sleep staging classification accuracy performance for five sleep stages classification task (CT-5). The overall classification accuracy reported as 99.34%, 90.8%, and 98.50% for SG-I, SG-II, and SG-III dataset, respectively.

Keywords Sleep stage · EEG signal · Feature screening · Machine learning algorithm · Stacking model

1 Introduction

Each and individual human needs to have proper sleep at night and it is the basic requirement for the human being and proper sleep quality plays a direct role in our day-to-day life. Sometimes its impact is reflected in our physiological activities such as the quality of learning ability, physical activity, mental ability, and performance of the overall activities (Aboalayon et al. 2016). In general

complete sleep, duration is covering through sleep stages at regular interval time and which is related to our brain neuron system (Chung et al. 2009). With the modern digital generation, the lifestyle of the human being is complicated and ultimately it has resulted that millions of people get a poor quality of sleep during night time. This problem is seen across the world with all age groups of people, and this is a global challenge in the health care sector because it has found from the different research that poor quality of sleep is the major responsibility of creating critical diseases such as bruxism (Heyat et al. 2019), insomnia (Lai et al. 2019), narcolepsy (Rahman et al. 2016), obstructive sleep apnea (Kim et al. 2018) and rapid eye movement behavioral disorder (Siddiqui et al. 2016). Finally, it creates

✉ Santosh Kumar Satapathy
santosh.satapathy@pec.edu

¹ Pondicherry Engineering College Puducherry, Puducherry, India

damages to various parts of our body such as heart failure, brain stroke, and several neurological disorders. Normally, sleep behavior is characterized by changes in activities in our body such as respiration and heartbeat rate, brain-behavior, and muscle movements. Currently, there are two important sleep standards followed during sleep staging analysis. According to both standards the whole sleep stages are divided into three basic categories: (1) wakefulness (W), (2) rapid eye movement (REM), (3) non-REM. The first sleep handbook was edited by Rechtschaffen and Kales (R&K) in 1968 (Rechtschaffen 1968) and under this sleep scoring rule, the NREM sleep stage is further divided into four sub-sleep stages: S1, S2, S3, and S4. These sleep standards followed all clinicians during the analysis of sleep irregularities of the patients, but in the year 2008, another recognized sleep recommendation was proposed by the American Academy of Sleep Medicine (AASM) (Iber 2007) and defined a new sleep handbook through small modification on the R&K rules. As per AASM rules, the non-rapid eye movement is further segmented into three sub-sleep stages such as N-REM1, N-REM2, and N-REM3 (Iber 2007). The sleep cycle generally repeated at regular intervals of time between NREM stages to REM stages and each duration of the sleep cycle is around 90–110 min (Carskadon and Dement 2017). The quality of sleep ratio is individually different from person to person according to their age. Stage1 of N-REM sleep (N-REM1) is light, where the subject's eye movements are slow and the muscle movements slow, in the N-REM2 sleep stage, the eye movements completely stop and the response of the brain becomes slower. So both N-REM1 and N-REM2 sleep stages are categorized as light sleep stages. Similarly, the N-REM3 and N-REM4 are called deep sleep, in which no eye movements have occurred and some muscle movements are appeared (Nagabushanam et al. 2019). Finally, in the REM stage, breathing is increased and the physical movements have been seen rapidly. Sleep staging is normally examined using polysomnographic (PSG) recordings from the admitted subject in the clinic.

Basically during the PSG test, different physiological recordings are collected from the subjects to measure the sleep quality during nighttime. The major recordings included during the assessment of sleep quality are electroencephalogram (EEG), electromyogram (EMG), electrooculogram (EOG), and electrocardiogram (ECG) (Holland et al. 1974). Among these, major researchers the priority toward an analysis of sleep pattern abnormality through EEG signal because the EEG signal provides the

information on brain activities and behavior of subjects during sleep. It can help to characterize the behavior of whole sleep stages through different frequency ranges of EEG signal which were segmented into δ band (< 3.5 Hz), θ band (4–7 Hz), α (8–13 Hz), β (14–30 Hz) and γ (30–80 Hz), ultimately it can support during the demonstration of the characteristics of different stages of sleep (Acharya et al. 2015). To correctly predict the sleep behaviors of the subjects, most of the clinicians used electroencephalogram, which directly provides information about brain activities. EEG recordings, its hectic situation for sleep experts to monitoring within the 30 s framework and fix the labeling of sleep stages (O'Reilly and Nielsen 2014; Hassan and Bhuiyan 2016a). This approach consumed more time and required more manpower for hours of sleep recordings. To overcome the difficulty of the manual approach, nowadays automated sleep scoring system is obtained to analyze the recorded EEG signals (Lei et al. 2018; Sharma et al. 2017; Zhu et al. 2014; Ronzhina et al. 2012) with subject to treatments various types of sleep and neurological disorders in real-time diagnosis. Some of these sleep scoring systems are based on polysomnographic signals recordings (Sharma et al. 2019a; Spriggs 2014; Hassan and Bhuiyan 2016b).

1.1 Literature review

The main aims of the literature study are analysis the existing contributions with subject to different methodologies and models used in recent sleep staging studies. Various computational methodologies were proposed by researchers to support sleep experts for assisting sleep staging. Those carried steps were on the information extraction (Polysomnography channel selection), on the preprocessing (removing the data artifacts and data normalization), on the feature extraction step (transformation of linear and nonlinear features), on the feature selection technique (identifying the most relevant features) and finally on the classification algorithm. These entire literature reviews are divided into two sections. In the first section, we discussed the different types of features extracted, and in the second part, we briefly discuss the different classification models used during sleep staging, which used recently sleep staging studies proposed for characterization the sleep-related abnormalities under sleep standards recommended by R&K and AASM manuals. Here we have presented some comparative studies regarding sleep staging.

Different Signal segmentation and feature extraction techniques used during Sleep staging

Zhu et al. (2014) proposed graph-based features for the analysis of the sleep patterns from EEG signals. The required sleep recordings were obtained from eight subjects having 14,963 epochs and the proposed SVM model achieved 87.5% classification accuracy.

Hassan and Bhuiyan (2016b) obtained wavelet transform concepts for decomposing the signal into different signal sub-bands using tunable Q-factor and extracted the spectral features. The entire experimental work conducted using 15,188 epochs from 28 subjects.

In Hassan and Bhuiyan (2016a), the author has introduced an empirical mode of decomposition for signal decomposition and extracted high order statistical features for sleep staging analysis.

Sharma et al. (2019a) used wavelet function and IIR filters for signal segmentation and artifact removals. To analysis, the sleep patterns of the subjects, three different categories of features were extracted such as time and frequency domain features and non-linear features.

Sharma et al. (2017) introduced the iterative filtering techniques for EEG signal segmentation. The obtained amplitude and frequency modulated components from iterative filtering decomposition used to extract the higher-order statistical features for discriminating the sleep characteristics.

Sen et al. (2014) proposed discrete wavelet transform techniques for signal enhancement and extracted four different categories of features such as time, nonlinear, frequency, and entropy-based features. The entire study was executed from 5160 epochs of 30 s length.

Memar and Faradji (2018) extracted the time-domain, non-linear, and entropy features from eight EEG signal sub-bands such as alpha, delta, theta, gamma, beta1, beta2, gamma1, and gamma2. The total study was conducted with 142,391 epochs of three different sleep datasets.

Tian et al. (2017) extracted multi-scale entropy properties from EEG signal for characterizing the signal in multiple temporal scales manner. The study was implemented with total epochs of 18,248 of 30 s length from 10 sleep disordered and 10 healthy subjects.

Alickovic and Subasi (2018) used multi-scale principal component analysis and the informative features are extracted from signal sub-bands using discrete wavelet transform techniques. Twenty subjects participated in this ensembling sleep staging analysis.

Different classification models obtained during sleep staging

Obayya and Abou-Chadi (2014) considered single-channel EEG signal as input to identifying sleep disorders and selected subjects for this experiment work limited between 35 and 50. Here authors have obtained wavelet concept techniques for feature extraction and classified the selected features using the fuzzy algorithm. The classification model provided 85% accuracy.

Zhu et al. (2014) introduced sleep analysis through horizontal and difference visibility graphs from input signals and finally, extracted properties from the input signals are forwarded to SVM classifiers for classifying multiple stages of sleep stages and their final accuracy was 87.50% for two-state sleep stage classification problems.

Diykh et al. (2016) introduced the concept of structural graph similarity and the experimental work completely based on EEG signals. The model obtained a classification accuracy of 95.93% using the SVM classifier.

Sriraam et al. (2016) obtained multiple-channel of EEG signals from ten healthy subjects. The extracted features processed through a multilayer perceptron feedforward neural network and the overall accuracy with 20 hidden units were reported as 92.9% and subsequently for 40,60,80 and 100 hidden units in MLP, it was reported as 94.6,97.2,98.8, and 99.2, respectively.

Silveira et al. (2016) used Discrete Wavelet Transform (DWT) for signal segmentation. The extracted features were applied to a random forest classifier and overall accuracy was reported as 90%.

In Hassan and Subasi (2017) the author obtained the bootstrap aggregating concept for classifying the sleep stages. The proposed model obtained two public sleep datasets such as the PhysioNet Sleep-EDF dataset and Dreams Subjects dataset and the model reported accuracy of 92.43% for two-class sleep stages classification.

Gunnarsdottir et al. (2018) have designed an automated sleep stage scoring system with overnight PSG data and the extracted properties were classified through DT classifiers. The overall accuracy for test set data was reported as 80.70%.

Memar and Faradji (2018) considered 25 sleep suspected subjects and 20 healthy subjects for experimental purposes. The extracted features validated through the Kruskal–Wallis test and applied random forest classifier and achieved an overall accuracy of 95.31%.

In Braun et al. (2018) the author proposed a portable and effective sleep scoring system, in which he has experi-

mented on the combination of features extracted from EEG signal and classifiers. He designed the system in such a manner that, the proposed research achieved the best classification accuracy by considering fewer frequency domain features, and the overall accuracy reported as 97.1% for the two-state classification problem.

In Huang et al. (2019), the author considered combinations of multi-variate signals for sleep scoring and extracted multiple features from both time and frequency domain features. The selected features are forwarded into SVM classifier and the reported accuracy with input single-channel EEG is 92.04%.

1.2 Contribution

It has been observed from the literature review that most of the researchers focused on accurate identification of different deployed sleep staging algorithms (O'Reilly and Nielsen 2014; Sharma et al. 2019b; Memar and Faradji 2018; Tian et al. 2017; Huang et al. 2019). The major influence to achieve high accuracy on sleep staging is selecting suitable feature parameters. It has been noticed that many of the papers did not specify this subject in detail, most of the sleep staging studies are based on signal characteristics parameters. It has been found that most of the sleep studies have been executed based on only one health conditioned subject so that sometimes it has been seen that the model may not be perfect for different categories of subjects. All these above-mentioned challenges we have addressed in our proposed sleep staging study.

In the present research on sleep staging, we propose an artificial intelligence-based automated sleep staging system using EEG signals from the subject with different health conditions subjects. The main important part of this work is the designed ensemble learning stacking model algorithm for improvement on the sleep staging classification accuracy for the five-to-two sleep states classification tasks. The proposed research completed majorly in five stages such as signal pre-processing, feature extraction, feature screening, classification algorithm, and finally the comparative analysis in between obtained different categories of subjects. The novelty of the proposed is the analysis of the sleep behavior of the subjects through proper feature screening techniques, which determines the statistically significant difference among the sleep states. Those features are selected further in the classi-

fication task, which is successfully screened through obtained feature selection techniques. The most important part of this research contribution is to propose a new ensembling learning classification model, called a stacking model for sleep staging analysis. The performance of the proposed methodology performed well for the multi-class sleep stages classification task incomparable to the existing sleep studies related to various performance evaluation metrics.

1.3 Structure of the paper

Section 2 presents briefly the proposed methodology including experimental data preparation, data preprocessing, feature extraction, and feature screening. In Sect. 3, we describe the classification algorithms used in this paper for sleep staging evaluation. Section 4 discusses the obtained experimental results from the proposed methodology from three subgroups of subjects' sleep recordings. In Sect. 5, we briefly discuss our proposed methodology results, advantages, and limitations and make a result analysis with the existing the published methods. Section 6 ends with concluding remarks with future work descriptions.

2 Materials and methodology

In this part, we presents the complete layout of the proposed multiple sleep staging classification tasks using ensemble learning stacking model. The main idea of this research work is to improvement on sleep staging accuracy with input of suitable screened features for multiple-class sleep states classification. The complete flowchart of this proposed methodology is shown in Fig. 1.

The complete and integrated implementation process of the proposed ensemble learning stacking model (Fig. 1) is described in Algorithm 1. In this algorithm, we briefly presents the individual procedure for each experimental phases such as (1) data preparation, (2) data pre-processing, (3) feature extraction, (4) feature selection, and finally (5) proposed ensemble learning stacking model.

Algorithm 1: Ensemble learning Stacking Model Algorithm**Input:** Three independent subgroups of sleep dataset are considered for input from a ISRUC-Sleep dataset**Output:** Final predictions from the meta-layer of stacking model**# Data Preparation****Input:** Raw Sleep_Dataset**Output:** Prepared Sleep_Dataset

// Split Sleep dataset according to the different session recordings

Procedure Split_Dataset (Sleep_Dataset)

```

For each Samples in Sleep_Dataset do
  IF (Subject condition=sleep disorder and session recordings=one)
    Create file called as SG-1
  End IF
  IF (Subject condition=sleep disorder and session recordings=two)
    Create file called as SG-2
  End IF
  IF (Subject condition=Healthy controlled and session_recordings=one)
    Create file called as SG-3
  End IF
End For
// Treatment of different missing session recordings
For each row in Prepared Sleep_Dataset
  For each column in Prepared Sleep_Dataset
    If column value=NAN
      Delete the row and its corresponding column
    End if
  End for
End for
Return (Prepared Sleep_Dataset)

```

End Procedure**#Data Pre-processing****Input:** Raw Sleep recordings from different medical conditioned sleep disordered and healthy subjects from the different subgroups of the sleep dataset**Output:** Filtered Sleep_Recordings**Initialization:** FS: Sampling frequency, NE: Number of epochs, SE: Samples per epoch, SL: Study length, FD=Filtered data, SS=Sleep stages, CE=Current epoch**Procedure** Preprocessing (Prepared Sleep_Dataset)

```

For all the subjects from the Prepared Sleep_Dataset do
  Extract the C3-A2 of EEG signal
  FD=wiener filter (EEGC3-A2@0.3-35Hz)
  SS=Sleep experts scores
  NE=SL (secs) ÷ 30 seconds/epoch
  SE=FS × 30
  For i=1 to NE do
    Get CE using SE and i
    [EEGC3-A2, SS]=FD (CE)
  End For
End For
Return (FD)

```

End Procedure**#Feature Extraction****Input:** Filtered data**Output:** Normalized Feature Vector**Initialization:** FV: Feature Vector, NFV: Normalize Feature Vector**Procedure** Feature_Extraction (FD)

```

For each row in FD do
  For each column in FD do
    Get time domain and frequency domain features
    FV={F1, F2, F3, ... .. FN}
  End For
End For
//Apply Normalization Techniques
For each row in FV do
  For each column in FV do
    Compute Min-Max normalization vector
    NFV={F1, F2, F3, ... .. FN}
  End For
End For
Return (NFV)

```

End Procedure

#Feature Selection

Input: Set of training instances

Output: Weight vector of the feature scores

Initialization: TD =Training instances, n =Total number of training instances, k =Number of features, m =training samples from the total training samples

Procedure Feature_Selection (TD)

```

Set  $WT[AT] = 0$ 
For  $i=1$  to  $m$  do
    Select the target instances  $B_i$  randomly
    Find the nearest hit ( $H$ ) and nearest miss ( $M$ ) from the instances
    For  $AT=1$  to  $k$  do
         $WT[AT] = WT[AT] - \text{diff}(AT_k, B_k, H) / M + \text{diff}(AT_k, B_k, M) / M$ 
    End for
End for
Return ( $WT[AT]$ )
    
```

End Procedure

#Ensemble learning Stacking Model

Input: $TD=\{X_i, Y_i\}$

Output: Predictions from the Ensemble Stacking Model

Initialization: TD =Training dataset, PD =Predictions of the base-layers models

Procedure Ensemble_Stacking_Model (TD)

```

Obtain cross validation approach for preparing the training dataset for meta-classifier
Split the  $TD$  in to  $P$  number of equal size subsets
 $TD=\{TD_1, TD_2, TD_3 \dots \dots \dots TD_P\}$ 
For  $P=1$  to  $M$  do
    Learn base layer classification namely SVM, DT, KNN, and RF
    For  $Q=1$  to  $N$  do
        Learn a classifier  $PD_{MN}$  from  $TD$ 
    End For
    Preparing a training data for meta-classifier (XGBoost)
    For each  $X_i \in TD_P$  do
        Train meta-classifier on new instances ( $X_i', Y_i$ )
        %  $X_i' = \{PD_{M1}(X_i), PD_{M2}(X_i), PD_{M3}(X_i) \dots \dots \dots PD_{MN}(X_i)\}$ 
    End For
End For
Return  $Y_i = \{Y_1, Y_2, Y_3 \dots \dots \dots Y_N\}$  from meta-classifier
    
```

End Procedure

2.1 Experimental data

In this study, the authors use three categories of subjects with different medical conditions. The recorded data were collected from the ISRUS-Sleep dataset, which was specifically for sleep studies. The data collection has been done at the Sleep Medicine Centre of the Hospital of Coimbra University (CHUC) (Khalighi et al. 2016). The first subsection includes 100 subjects, the second subsection includes 8 subjects, and finally, the third subsection includes information from 10 healthy subjects' sleep recordings. The data acquisition process is conducted in an 8–9 h full night PSG test at healthcare facilities. Signals

are sampled at 200 Hz, and each epoch length is considered a timeframe of 30 s according to the AASM standard.

In this study, the authors used the C3–A2 channel for computing sleep stage classification. We considered 4 subjects one session sleep recordings from subgroup-I (SG1), 4 subject's two-session recordings from subgroup-II (SG2), and 4 subjects from subgroup-III (SG3) with one session recordings. The SG1 (Subgroup-I) consists of 3000 epochs, while SG2 (Subgroup-II) contained 6000 epochs, and finally SG3 (Subgroup-III) consists of 3000 epoch and each epochs size is 30 s length. The entire sleep staging process was executed with all these epochs according to AASM standards. The numbers of epochs distributed

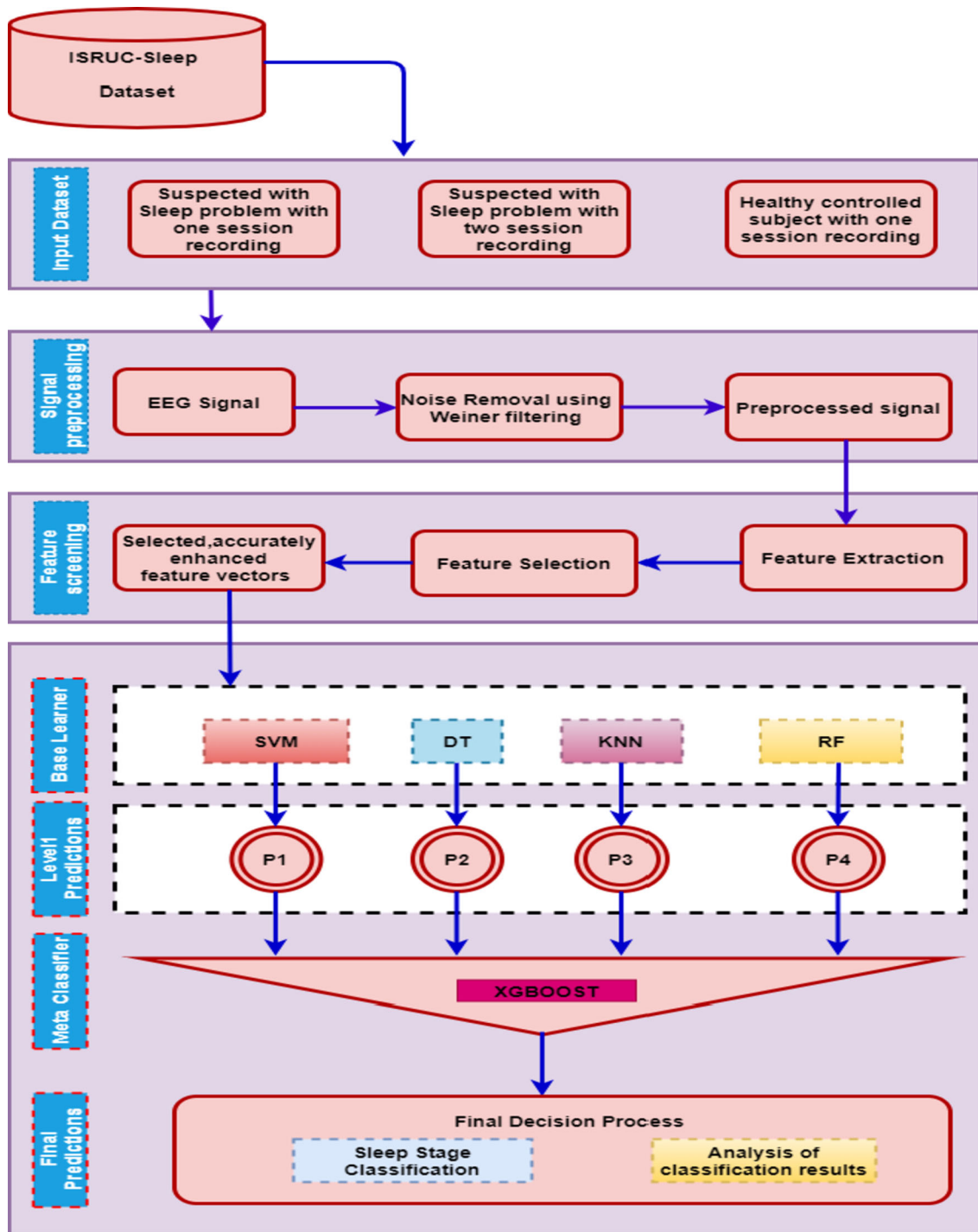


Fig. 1 Proposed flowchart of multiple sleep staging classification tasks using ensemble learning stacking model

among into the different sleep stages are presented in Table 1.

2.2 Signal preprocessing

This process starts with the segmentation of the raw signal from the three different medical conditioned subjects from

the ISRUC-Sleep dataset. Each epoch of length is the 30 s and each 30 s epochs contains 6000 sample points, as the sampling frequency is 200 Hz. As we have already discussed that sleep recordings are highly complicated and random, sometimes it has found that some of the irrelevant signal information is overlapped with the brain EEG signals. Majorly, the artifacts like eye blinks and muscle

movements signals are composition within the originally recorded signals. These contaminated signals give inaccurate analysis toward sleep staging classifications. To remove these artifacts from the original sleep recorded signals, we have proposed wiener filter techniques (Borowicz 2017). It is one of the parametric techniques, which supports removing the unwanted forms of noises from the original brain EEG signal. This filter mechanism is completely based on stational approaches, and it minimizes the mean square error between the estimated recording signals and the desired signals.

2.3 Feature extraction

It is one of the important step during the automated sleep staging analysis based on the EEG signals because the recorded signals are the compositions of various signal sub-bands. To analysis the changes sleep behavior from each sub-bands, extracting the features has been the best way toward recognizing the changes sleep characteristics over the individual sleep stages. The extracted signal properties may easier during discrimination the sleep patterns and alternatively it also put impacts upon the sleep staging classification performances. In the recent research progress, most of the contributions majorly focused on the three different categories of feature for signal analysis such as time domain, frequency domain, and nonlinear features (Yasoda et al. 2020; Ahmad et al. 2018; Gupta et al. 2015; Al Ghayab et al. 2019; Al-Janabi et al. 2019; Souri et al. 2020; Schetinin and Schult 2006). In this work, we obtain both time and frequency domain features. 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) alpha (α) rhythm, sigma (σ) rhythm, and beta (β) rhythm patterns in the EEG signals in a different frequency ranges. In this proposed research work, we extracted 28 features which include (1) 13 time-domain based features, (2) 15 frequency domain based features.

The extracted different features in this proposed study and their descriptions are summarized in Table 2, was performed an epochs of 30-s length of the filtered EEG signals.

After extraction of features from respective input channels, we have applied normalization techniques before being obtained to feature selection techniques, and here we use to zero mean and unit variance, and a normalized feature vector is generated. Generally, it supports to increase the performance of the system.

2.4 Feature screening algorithms

Next to feature extraction, it is also important to select the suitable features because it has found that sometimes all the extracted features may not be relevant for all the subject cases and alternatively it directly also affects the performance of the classification models. The main intention behind considering the feature screening is to find out the most commendable features which help to discriminate the changes in sleep characteristics in the different sleep stages. Here we obtained the ReliefF feature selection algorithm for identifying the relevant features. It is one of the individual feature filtering selection methods, in which the quality and relevance of the features to be measured (Robnik-Šikonja and Kononenko 2003). The essential concept behind this algorithm is to select highly commendable features which help to discriminate the sleep behavior of the subject. As an output, this algorithm assigned a weight to individual input features according to their relevance. It determines how far the features are most discriminate to different instances amounts to different sleep stages. It generates a weight for each feature which is in the ranges from -1 (worst) to $+1$ (best) score. The larger the weight of the feature, the higher the association between the features and sleep stages. The main advantage of this algorithm is well managed with noisy and unknown data.

2.5 Proposed ensemble stacking model

Ensemble techniques are one of the ML approaches for improving the model by combining several models (He

Table 1 Number of 30 s epochs from each sleep stage of ISRUC-sleep dataset records

Sleep stages	No. of SG1 epochs	No. of SG2 epochs	No. of SG3 epochs
Wake	657	1305	648
N1	471	1122	349
N2	1064	1710	942
N3	485	1038	674
REM	323	825	387
Total	3000	6000	3000

Table 2 Short explanation of the extracted features for this proposed study

Time-domain based features		
Feature No.	Extracted feature equation	Feature descriptions
1	$\text{Mean}(\bar{x}) = \frac{1}{N} \sum_{i=1}^N x_i$ <p>with N = the length of the data sample x and \bar{x} is the mean of the data sample</p>	The mean electrical potential of an epoch is calculated. It also measures the central tendency in the data points
2	$\text{Maximum} = \text{Max}[x_i]$	It is used to quantify the range of data and it helps to find the magnitude of signal baseline
3	$\text{Minimum} = \text{min}[x_i]$	
4	$\text{Variance (Var)} = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}$	It helps to determine how the data is dispersion with respect to the value of the mean
5	$\text{Standrad deviation (SD)} = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^2$	It is used to calculate the quantity of variation and dispersion of the data
6	$\text{Median} = \left(\frac{N+1}{2} \right)^{\text{th}}$	It helps to get the information about the center and spread of the signal data
7	$75\text{th percentile}(75^{\text{th}}_p) = \text{Max}(x_i) < P\{75\}$ <p>x_i is the signal data $p\{75\}$ = the 75th percentile of the signal data</p>	The percentile analysis provides amplitude information of the signal, which helps to discriminating the sleep stages. It defines the value below which 75% of the random variables values data is located
8	$\text{Signal Skewness (skew)} = \sum_{i=1}^N \frac{(E(x_i - \bar{x})^3)}{\sigma^3}$ <p>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</p>	The skewness helps to measure the symmetry of the signals distribution with respect to the mean value. The normal distribution of the signal is zero, while the positive and negative skewness indicates that the data are skewed in to the right and left hand sides. It is a higher-order statistic measure (third moment)
9	$\text{Signal kurtosis (kurt)} = \sum_{i=1}^N \frac{(E(x_i - \bar{x})^4)}{\sigma^4}$ <p>where σ denotes as standard deviation of the sample data for all i, and E is the expected mean value</p>	It measures whether the data is peaked or flat relative to the normal distribution. It is a higher-order statistic measure (fourth moment)
10	$\text{Signal Activity} = \text{Var}(x_i)$	Among the time-domain features, Hjorth parameters (Activity, Mobility, and Complexity) are more popular for interpreting the EEG signals, which directly helps during the sleep stages classification. 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	$\text{Signal Mobility} = \sqrt{\text{Var}(x'_i) / \text{Var}(x_i)}$	
12	$\text{Signal Complexity} = \sqrt{\text{Var}(x''_i) X \text{Var}(x_i) / \text{Var}(x'_i)^2}$	

Table 2 (continued)

Time-domain based features		
Feature No.	Extracted feature equation	Feature descriptions
13	Zercrossing Rate(ZC) $= \sum_{i=1}^N ZC'(i)$ $ZC'(i) = \begin{cases} 1, & x(i) \leq 0 \cap x(i+1) \\ 1, & x(i) \geq 0 \cap x(i+1) \\ 0, & \text{otherwise} \end{cases}$	It provides information with regards to the number of instances where the EEG signal crosses the references line
<i>Frequency domain-based features</i>		
14, 15, 16, 17	Relative Spectral ower ($\delta, \theta, \alpha, \beta$) $= \frac{\int_{-f_1}^{f_0} x_i(f)df + \int_{f_0}^{f_1} x_i(f)df}{\int_{-\infty}^{\infty} x_i(f)df}$	It is one of the popular parameter 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 β frequency sub-bands
18, 19, 20, 21, 22, 23, 24	Power Ratio ($\delta, \theta, \alpha, \beta$)	It is used for computing the power from the current and background of the epochs with the different frequency ranges for interpreting the behavior of the sleep stages
25, 26, 27, 28	Power spectrum ($\alpha, \beta, \theta, \delta$) $= \sum_{i=1}^N x_i$	It helps to retrieve the information on how the intensity of a time-series signal data is distributed in the frequency domain

et al. 2019). The main advantage of this work is to reduce the variance and bias factor with the proposed stacking algorithm. It helps to increase the accuracy of the model and reduce the variability of the prediction. The proposed ensemble learning stacking model, which is in form of two-layer architecture. The first layer data and model to be treated as base-layer data and models, respectively, similarly, the second layer cross-validated data and models to be considered as meta-layer data and model, respectively. It adopts the parallel structure since the base-layer models are trained independently from each other. The base-layer predictions become input for the meta-layer, and finally, the meta classifier delivers the final predictions.

The most important advantage of the proposed model can analyze the complex sleep patterns in the meta-layer incomparable to the other parallel ensemble learning approaches. Additionally meta-layer also helps to recognize the consistently mis-predicts samples from the feature space of the base-layer classifiers due to incorrect learning and select the suitable base-layer models which are more suitable for that specific feature space. The same problems may not possible using bagging and boosting techniques. In this proposed stacking algorithm, we also use the cross-validation techniques for preventing the information leakage from the base-layer models to the meta-layer model.

3 Experimental results

The entire experiment of this research work is coded and implemented carried using the MATLAB R2017a software for signal preprocessing, feature extraction, and feature screening and also we obtained the Scikit-machine learning tool for XGBoost and mlxtend library for stacking algorithms [97] in Python running on a personal laptop with an Intel Core™ i3-4005U CPU 1.70 GHz, 2 core(s), 4 logical processors, 4 GB RAM and Windows 10 operating system. To evaluate the performance of the proposed method, a set of experiments were conducted using subjects with three different categories of medical conditions. The first experiment was conducted with the subject who was affected with different types of sleep problems and their one-session recordings considered the experiment, the second experiment carried with subjects having sleep problems and includes two different session recordings, which were recorded on two different dates. Finally, the third experiment was conducted with subjects who were completely healthy and no prior types of diseases related to sleep. In this research work, we have only considered single-channel of EEG signal as C3–A2 for extracting the sleep behavior of the subjects, according to recent sleep staging, it has found that the C3–A2 channel is most effective in analyzing the brain behavior for sleep studies in terms of classification accuracy performances because it provides central part of the brain information with related to analysis the brain-behavior during sleep (O'Reilly and Nielsen 2014; Hassan and Bhuiyan 2016a, b; Lei et al.

2018; Sharma et al. 2017, 2019a, b; Zhu et al. 2014; Ronzhina et al. 2012; Spriggs 2014; Şen et al. 2014; Memar and Faradji 2018; Tian et al. 2017; Alickovic and Subasi 2018). Hence, we have used the C3–A2 channel of EEG signal recordings for sleep staging for all the three categories of subject data in our proposed work.

In this research work, we have conducted four individual experiments for sleep staging. The first three experiments were conducted with three different subgroups subjects' sleep recordings with help of base layers classification algorithms such as SVM, DT, KNN, and RF. The final experiment was our proposed ensemble learning stacking model, where we used another layer for learning the model, called as meta classification layer. We used the XGBoosting algorithm in the Meta classifier layer, which considered input as base layers predictions and made final decisions on sleep scoring. In the proposed sleep study, we conducted multi-class sleep stages classification: CT-2 (Wake vs. Sleep), CT-3 (Wake vs. NREM vs. REM), CT-4 (Wake vs. N1 + N2 vs. N3 vs. REM) and CT-5 (Wake vs. N1 vs. N2 vs. N3 vs. REM).

3.1 System performance evaluation metrics

The proposed study provides an in-depth analysis based on a comparative analysis of multiple different subjects with different session recordings obtained for sleep stage scoring analysis. For that reason, the authors have used multiple evaluation metrics to analyze the performance of the proposed sleep stage classification method.

This study considers four criteria such as the classification accuracy (Raza et al. 2016), recall (Bajaj and Pachori 2013), specificity (Hsu et al. 2013), precision (Zibrandtsen et al. 2016), F1Score (Berry et al. 2014), and kappa score (Statistic and in Reliability Studies: Use, Interpretation, and Sample Size Requirements 2005). The confusion matrix is used to evaluate the results obtained from classification algorithms. The authors have analyzed the information about the actual and predict score achieved by the algorithms used.

3.2 Performance of sleep staging using Subgroup-I (SG-I) dataset

In this experiment, we considered with single-channel C3–A2 of EEG signal recordings of four sleep-disordered subjects. All four subjects have diagnosis with sleep syndrome disease along some of them also suffered from insomnia and other types of sleep problems. The selected features are forwarded into the obtained four base layers classifiers such as SVM, DT, KNN, and RF for initial predictions on sleep staging. The classification accuracies (CAs) achieved for the two-five sleep states from base

classifiers are observed. Table 3 presents the summary of the reported classification accuracy results (CAs in %) from obtained classifiers.

It has observed from Table 3 that, the highest classification accuracy reported for CT-2 (99%), CT-3 (97.97%), CT-4 (97.77%) and CT-5 (96.33%) through RF classifier, similarly the lowest accuracy for CT-2 (93.87%), CT-3 (89.27%), CT-4 (85.07%) and CT-5 (84%) noticed from DT classifier. The highest accuracy reported RF classifier confusion matrix results for CT-5 is presented in Table 4.

3.3 Performance of sleep staging using Subgroup-II (SG-II) dataset

In this section, we have considered four subjects and their two different session sleep recordings, which were recorded on two different dates from the enrolled subjects. The subjects were under-diagnosis of roncopatia disease and also some of them were affected with various types of sleep problems. The same channel was used for signal acquisition and extracted the same features from the processed EEG signal. The same obtained classification algorithms used here and the reported classification accuracy performance from all the classifiers are shown in Table 5.

From classification performance results, it has been found that the RF classification model performs well for two-class to five-class classification tasks. The reported accuracy for CT-2 (98.62%), CT-3 (90.2%), CT-4 (89.71%) and CT-5 (89.1%). The KNN model achieved very low accuracy in terms to sleep scoring for CT-2 (90.33%), CT-3 (88.77%), CT-4 (84.54%), and CT-5 (82.15%). The confusion matrix for CT-5 using RF classifier is presented in Table 6.

3.4 Performance of sleep staging using Subgroup-III (SG-III) dataset

In this experiment, we have obtained completely four healthy controlled subjects' sleep recordings, who have not affected any types of sleep problems during sleep hours.

Table 3 Summary of the accuracy results for different classifiers with SG-I dataset

Classifier	CT-2 (%)	CT-3 (%)	CT-4 (%)	CT-5 (%)
Quadratic SVM	96.70	92.30	90.10	88.90
Gaussian SVM	97.40	93	91.30	89.80
DT	93.87	89.27	85.07	84.00
KNN	96.02	91	88.90	87.50
RF	99.00	97.97	97.77	96.33

Table 4 Confusion matrix for CT-5 using Random forest classifier using SG-I dataset

True/predicated	Wake (0)	N1 (1)	N2 (2)	N3 (3)	REM (5)
Wake(0)	601	21	20	10	5
N1(1)	6	458	4	0	3
N2(2)	7	1	1052	3	1
N3(3)	4	1	7	473	0
REM(5)	0	2	5	1	315

The performance achieved from base learning classifiers using the SG-III dataset is presented in Table 7.

It has observed from Table 7 that, the highest classification accuracy performance reached with random forest classification model for CT-2 (99.10%), CT-3 (98.07%), CT-4 (97.93%), and CT-5 (98.13%). The confusion matrix for reporting higher accuracy with RF model is presented in Table 8 with respect to CT-5.

3.5 Performance of sleep staging using proposed Ensemble stacking algorithm using subgroup-I/II/III datasets

Finally, we have deployed our proposed ensemble techniques for sleep scoring, using the integration of the base layers classification model. In this study, we have used four classification techniques such as SVM, DT, KNN, and RF for taking the first predictions on sleep scoring. This computed first layer prediction output is considered as to the second layer, called as meta classification layer. We obtained an XGBoosting classification algorithm in the second layer, which integrates the previous layer's predictions and generates the final decisions on sleep staging. The confusion matrix of this proposed ensemble learning stacking model using SG-I, SG-II, and SG-III datasets for five sleep states classification problems are presented in Tables 9, 10, and 11, respectively. The final classification performance of this proposed model with input of SG-I, SG-II, and SG-III datasets are presented in Table 12.

It has found from Table 12 that, the proposed ensemble learning stacking model performed excellent with related to five sleep states classification problem in comparable to base layers classification algorithms such as SVM, KNN, DT and RF. The proposed stacking model reported classification accuracy of 99.34%, 96.30%, and 98.50% for SG-I, SG-II, and SG-III datasets, respectively.

3.6 Summary of results

In this paper, we were used three subgroups data of ISRUC-Sleep for sleep staging. In all the experimental part

Table 5 Summary of the accuracy results for different classifiers with SG-II dataset

Classifier	CT-2 (%)	CT-3 (%)	CT-4 (%)	CT-5 (%)
Quadratic SVM	98.3	85.7	81.8	80.8
Gaussian SVM	98.6	86.66	83.31	81.27
DT	94.07	87.47	82.08	81.03
KNN	90.33	88.77	84.54	82.15
RF	98.62	90.2	89.71	89.1

Table 6 Confusion matrix for CT-5 using Random forest classifier using SG-II dataset

True/predicated	Wake (0)	N1 (1)	N2 (2)	N3 (3)	REM (5)
Wake(0)	1256	16	7	11	15
N1(1)	36	997	55	35	63
N2(2)	50	47	1508	40	78
N3(3)	14	35	42	885	7
REM(5)	26	30	40	7	700

Table 7 Summary of the accuracy results for different classifiers with SG-III dataset

Classifier	CT-2 (%)	CT-3 (%)	CT-4 (%)	CT-5 (%)
Quadratic SVM	98.62	95.19	92.69	91.12
Gaussian SVM	98.72	95.49	93.37	90.51
DT	94.47	89.53	83.97	80.02
KNN	87.71	86.64	83.23	80.59
RF	99.10	98.07	97.93	98.13

conducted for this study using single-channel EEG signal. The first three experiments conducted based on base layers learning classification model and the final experiment executed with the proposed stacking model. Table 13 presents the summary results that was obtained through different classification models of base layers and ensemble learning of stacking model with the three different health categories of subjects sleep recordings using SG-I, SG-II, and SG-III datasets.

The highest sleep staging was obtained using ensemble learning stacking model from all the three subgroups dataset. The highest classification accuracy was reported as 99.34% for five sleep states classification problem with SG-I dataset.

Table 8 Confusion matrix for CT-5 using Random forest classifier using SG-III dataset

True/Predicated	Wake (0)	N1 (1)	N2 (2)	N3 (3)	REM (5)
Wake(0)	614	0	1	4	0
N1(1)	4	380	4	1	0
N2(2)	3	3	1034	3	3
N3(3)	3	3	8	580	1
REM(5)	1	2	8	4	336

Table 9 Confusion matrix for CT-5 using proposed ensemble learning stacking model using SG-I dataset

True/Predicated	Wake (0)	N1 (1)	N2 (2)	N3 (3)	REM (5)
Wake(0)	643	1	10	2	1
N1(1)	6	458	4	0	3
N2(2)	7	1	1052	3	1
N3(3)	4	1	7	473	0
REM(5)	0	2	5	1	315

Table 10 Confusion matrix for CT-5 using proposed ensemble learning stacking model using SG-II dataset

True/Predicated	Wake (0)	N1 (1)	N2 (2)	N3 (3)	REM (5)
Wake(0)	1256	16	7	11	15
N1(1)	36	1062	30	35	23
N2(2)	50	47	1548	40	38
N3(3)	14	35	22	905	7
REM(5)	26	30	40	7	700

4 Discussion

A machine learning (ML) based automated sleep scoring is suggested for the classification of multiple sleep stages. Several experiments were executed on three subgroups datasets to validate the potency of the proposed methodology. The proposed sleep staging methodology can automatically learn high level from the single-channel EEG signal directly.

4.1 Hypothesis and limitations of the proposed model

The proposed automated sleep staging results indicate that the ensemble learning stacking model reported improvement in classification accuracy compared to the other

Table 11 Confusion matrix for CT-5 using proposed ensemble learning stacking model using SG-III dataset

True/Predicated	Wake (0)	N1 (1)	N2 (2)	N3 (3)	REM (5)
Wake(0)	615	0	1	3	0
N1(1)	4	381	3	1	0
N2(2)	3	3	1034	3	3
N3(3)	3	3	3	585	1
REM(5)	1	2	4	4	340

earlier automated sleep staging and other ML methods. In comparison with the other contributed sleep staging classification problems, we mainly focus on four things that help to improve the classification accuracy of sleep staging: 1. preprocessing the signal using wiener filter techniques, 2. feature screening algorithm using ReliefF weight techniques, and 3. ensemble learning classification techniques.

One of the major advantages of this proposed model is using three different medical conditioned subjects with their different sleep recordings. One of the other major focused things in this work was the proper screening of the features by considering the ReliefF weight algorithm and eliminating redundant features using Pearson correlation techniques. In this research work, we have used wiener filtering techniques which support a lot toward reducing the artifacts from the input channel. Our proposed methodology was executed with 12,000 epochs, each epoch of 30 s length for two-five sleep stages classification with three subgroups datasets.

The major difficulty with PSG signal is made the more uncomfortable situation for subjects during sleep recordings due to more connecting electrodes and rigid placement postures (Al-Janabi et al. 2019). Sometimes this uncomfortable scenario and possible body movements during sleep recordings degrade the quality of input signals. Finally, we have obtained the ensemble learning model which combinations of two-layers of classification models, first layer is considered as the base learning layer, and the second layer is called a meta-learning layer. The final decisions to be obtained from the meta-layer classification model. Even though our proposed framework achieved very good accuracy incomparable to other state-of-the-art works. Till, there is more way for us to improve our model to reach classification accuracy closer to 100% and this result we can achieve by overcoming the class imbalance problem. Also, we will use deep learning techniques, in which automated feature learning helps us to overcome variations on feature influence problems and supports to reach the high classification accuracy on sleep staging. The class imbalance problem solves by implementing data augmentation concepts using deep learning techniques.

Table 12 Performance of five-state sleep (CT-5) staging based on ensemble learning stacking model with SG-I, SG-II, and SG-III datasets

Performance metrics	SG-I dataset (%)	SG-II dataset (%)	SG-III dataset (%)
Accuracy	99.34	96.30	98.50
Precision	98.59	90.91	98.43
Sensitivity	97.81	90.97	98.27
F1-Score	98.20	90.91	98.34

Table 13 Summary results of classification accuracy for various classification models and subgroups datasets

Database		Classification model	Accuracy rates (%) Five Sleep States (CT-5) Classification Task
ISRUC-Sleep	Subgroup-I	Quadratic SVM	88.90%
		Gaussian SVM	89.80%
		DT	84.00%
		KNN	87.50%
		RF	96.33%
		Stacking Model	99.34%
ISRUC-Sleep	Subgroup-II	Quadratic SVM	90.8%
		Gaussian SVM	91.27%
		DT	81.03%
		KNN	82.15%
		RF	89.1%
		Stacking Model	90.8%
ISRUC-Sleep	Subgroup-III	Quadratic SVM	91.12%
		Gaussian SVM	90.51%
		DT	80.02%
		KNN	80.59%
		RF	98.13%
		Stacking Model	98.50%

4.2 Comparisons of the proposed research work performances with other single-channel sleep staging classification

The authors have compared the proposed system performance with other studies available in state of the art. Therefore, the authors have selected studies with similar datasets according to our proposed study and based on a single channel. In Table 14, the features used in the proposed research work are compared to others used by related works using single-channel EEG signals of ISRUC-Sleep dataset. The comparisons with other similar research proposals available in the literature must take into consideration the use of single-channel EEG, different features and classification models is presented in Table 15.

Tables 14 and 15 present a comparison of the overall accuracies presented by the proposed methods and the available studies in the literature. The results reported, but the authors using the ensemble learning stacking model for

SG-III, SG-II, and SG-I present the highest results when compared with the other methods available in the literature. The proposed research work has obtained best accuracy results for five-sleep (CT-5) classification. We achieved an overall accuracy of 99.34% with SG-I, 90.8% with subgroup-II and 98.50% with SG-III subjects through stacking model. These classification results state that the proposed research work provides a significant contribution to the field of automated sleep stage classification and can support the decision-making by providing an efficient health-care model to diagnosis the sleep-related diseases.

5 Conclusion and future directions

This paper has presents an ensembling learning stacking model under AASM sleep scoring rules for two–five sleep states classification using a single channel EEG signals. The proposed methodology has been analyzed 12,000

Table 14 Comparisons of CAs (in %) of our proposed model with other state-art-of the techniques used same features and datasets

Author	Features	Classifier	Classification accuracy
Khalighi et al. (2011)	Statistical features, Entropy features, Time and Frequency domain features	SVM	95%
Simoes et al. (2010)	Harmonic and Hjorth parameters, Spectral Power, Skewness, Kurtosis	Bayesian Classifier	83%
Khalighi et al. (2016)	Statistical, Harmonic, Hjorth parameters, Entropy, Spectral power, Percentile (20,50,75)	SVM	93.97%
Sousa et al. (2015)	Skewness, kurtosis, Renyi entropy, Tsallis entropy, Shannon entropy, slow-wave index, autoregressive coef., peak to peak amplitude and Energy, Percent of energy, mean, standard deviation, relative spectral power, harmonic parameters, Hjorth parameters, percentile (20,50,75)	SVM	86.75%
Khalighi et al. (2013)	Time and Frequency domain features	SVM	81.74%
Tzimourta et al. (2018)	Energy	Random Forest	75.29%
Najdi et al. (2017)	Multi-domain features	Stacked Sparse Auto-Encoders (SSAE)	82.3%
Kalbkhani et al. (2018)	Entropy features	SVM	83.33%
Proposed	Mean, Minimum, Median, Zero Crossing Rate, Signal Skewness, Signal Activity, Signal Complexity, Signal Mobility, Maximum, Standard Deviation, Variance, 75 percentile, Signal Kurtosis, Relative Spectral Power, Band power, Power Ratios	Stacking Model (SG-I)	99.34%
		Stacking Model (SG-II)	90.8%
		Stacking Model (SG-III)	98.50%

epochs composed of three different medical conditioned subject datasets.

There are three main important contributions in this research study. The first one has obtained wiener filter techniques for eliminating the various artifacts that exist in the input signals and followed by we have extracted numerous features from linear and nonlinear features. This set of features supports the analysis of sleep EEG parameters and their characteristics. It has been observed that multi-feature extraction improves sleep staging accuracy. Secondly, the proposed research work obtained feature screening techniques, which directly useful for identifying the most relevant features from extracted feature vectors. Additionally, we also eliminate the redundant features from selected relevant features using Pearson correlation analysis, which helps us to select the suitable features for classification tasks. Thirdly, this proposed research work establishes an ensemble learning model, which integrates multiple classification models implements in two layers. In the first base layers, we obtain SVM, DT, KNN, and RF classifiers, and the second layer contained XGBoosting techniques. The proposed stacking model reported high

recognition rates for multiple sleep staging classification tasks.

The results show that this study can provide an effective mechanism for handling different health conditions of the subjects with high accuracy. The authors have also compared the proposed research with other similar studies available in the literature to show better performance and state the present contribution. Moreover, the proposed system presents high performance through all the categories of subjects with different medical conditions. Finally, this study uses the ISRUC-Sleep three subgroups dataset, which is a useful information source for a sleep evaluation.

Furthermore, the authors use shorter epochs lengths (the 20 s, 15 s, 10 s, and 5 s) to the development of a real automated sleep staging system. Consequently, the authors aim also to consider a high number of clinical sleep data. In particular, the future work will include the different sleep problems patients to evaluate the performance of the proposed method work for higher accuracy using deep learning techniques.

Table 15 Comparisons of sleep staging accuracy (%) between the proposed methodology and the recent methods

Author	Features	Classifier	Classification accuracy
Huang et al. (2019)	Time and frequency domain features	SVM	92.04%
Dhok et al. (2020)	Wigner–Ville Distribution(WVD) + Renyi Entropy features	Gaussian kernel-based support vector machine	87.45%
Wang et al. (2019)	Entropy and Energy, standard deviation, Hjorth parameters, Kurtosis, Skewness	Stacking Model	96.6
Mendez et al. (2016)	Mean, standard deviation, skewness, kurtosis and energy	k–Nearest neighbors classifier	80%
Hassan et al. (2015)	Spectral features	Boosted decision tree	82.03%
Sharma et al. (2019a)	Log-energy (LE), signal-fractal-dimensions (SFD), and signal-sample-entropy (SSE) from all seven signal sub-bands	SVM	91.5%
Hassan and Bhuiyan (2016a)	Statistical moment based features	Bagging	90.69%
Da Silveira et al. (2016)	Discrete wavelet transform	RF	91.5%
Memar and Faradji (2018)	Multiple features	RF	95.31%
Santaji and Desai (2020)	Mean, Entropy, Energy SiS, PSD	RF	97.8%
Proposed	Mean, Minimum, Median, Zero Crossing Rate, Signal Skewness, Signal Activity, Signal Complexity, Signal Mobility, Maximum, Standard Deviation, Variance, 75 percentile, Signal Kurtosis, Relative Spectral Power, Band power, Power Ratios	Stacking Model (SG-I)	99.34%
		Stacking Model (SG-II)	90.8%
		Stacking Model (SG-III)	98.50%

Funding This study is not funded by any organization.

Data availability All EEG files are available from the ISRUC-SLEEP database (https://sleeptight.isr.uc.pt/ISRUC_Sleep/).

Declarations

Conflict of interest All the authors in the paper have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

- Aboalayon K, Faezipour M, Almuhammadi W, Moslehpour S (2016) Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation. *Entropy* 18(9):272. <https://doi.org/10.3390/e18090272>
- Acharya UR et al (2015) Nonlinear dynamics measures for automated EEG-based sleep stage detection. *Eur Neurol* 74(5–6):268–287
- Ahmad W, Ahmad A, Lu C et al (2018) A novel hybrid decision support system for thyroid disease forecasting. *Soft Comput* 22:5377–5383. <https://doi.org/10.1007/s00500-018-3045-9>
- Al Ghayab HR, Li Y, Siuly S et al (2019) Epileptic seizures detection in EEGs blending frequency domain with information gain technique. *Soft Comput* 23:227–239. <https://doi.org/10.1007/s00500-018-3487-0>
- Alickovic E, Subasi A (2018) Ensemble SVM method for automatic sleep stage classification. *IEEE Trans Instrum Meas* 67(6):1258–1265. <https://doi.org/10.1109/TIM.2018.2799059>
- Al-Janabi S, Mohammad M, Al-Sultan A (2019) A new method for prediction of air pollution based on intelligent computation. *Soft Comput*. <https://doi.org/10.1007/s00500-019-04495-1>
- Bajaj V, Pachori RB (2013) Automatic classification of sleep stages based on the time-frequency image of EEG signals. *Comput Methods Programs Biomed* 112(3):320–328. <https://doi.org/10.1016/j.cmpb.2013.07.006>
- Berry RB, Brooks R, Gamaldo CE, Hardsim SM, Lloyd RM, Marcus CL, Vaughn BV (2014) The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. American Academy of Sleep Medicine
- Borowicz A (2017) Correction of eye-blink artifacts in EEG recordings using Wiener filtering. In: 2017 signal processing: algorithms, architectures, arrangements, and applications (SPA), Poznan, pp 172–177. <https://doi.org/10.23919/SPA.2017.8166859>

- Braun ET, Kozakevicius ADJ, Da Silveira TLT, Rodrigues CR, Baratto G (2018) Sleep stages classification using spectral based statistical moments as features. *Rev Inf Teórica Apl* 25(1):11. <https://doi.org/10.22456/2175-2745.74030>
- Carskadon MA, Dement WC (2017) Normal human sleep: an overview. In: Kryger M, Roth T, Dement WC (eds.), *Principles and Practice of sleep medicine*, 6th edn., Elsevier, Amsterdam, pp. 15–24. <https://doi.org/10.1016/B978-0-323-24288-2.00002-7>
- Chung MH, Kuo TB, Hsu N, Chu H, Chou KR, Yang CC (2009) Sleep and autonomic nervous system changes—enhanced cardiac sympathetic modulations during sleep in permanent night shift nurses. *Scand J Work Environ Health* 35(3):180–187. <https://doi.org/10.5271/sjweh.1324>
- Da Silveira TLT, Kozakevicius AJ, Rodrigues CR (2016) Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain. *Med Biol Eng Comput* 55(2):343–352. <https://doi.org/10.1007/s11517-016-1519-4>
- Dhok S, Pimpalkhute V, Chandurkar A, Bhurane AA, Sharma M, Acharya UR (2020) Automated phase classification in cyclic alternating patterns in sleep stages using Wigner-Ville distribution based features. *Comput Biol Med* 119:103691. <https://doi.org/10.1016/j.compbimed.2020.103691>
- Diykh M et al (2016) EEG sleep stages classification based on time domain features and structural graph similarity. *IEEE Trans Neural Syst Rehabil Eng* 24(11):1159–1168. <https://doi.org/10.1109/tnsre.2016.2552539>
- Gunnarsdottir KM, Gamaldo CE, Salas RME, Ewen JB, Allen RP, Sarma SV (2018) A novel sleep stage scoring system: combining expert-based rules with a decision tree classifier. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). <https://doi.org/10.1109/embc.2018.8513039>
- Gupta A, Agrawal RK, Kaur B (2015) Performance enhancement of mental task classification using EEG signal: a study of multivariate feature selection methods. *Soft Comput* 19:2799–2812. <https://doi.org/10.1007/s00500-014-1443-1>
- Hassan AR, Bhuiyan MIH (2016a) Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. *Biomed Signal Process Control* 24:1–10. <https://doi.org/10.1016/j.bspc.2015.09.002>
- Hassan AR, Bhuiyan MI (2016b) A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. *J Neurosci Methods* 15(271):107–118. <https://doi.org/10.1016/j.jneumeth.2016.07.012>
- Hassan AR, Subasi A (2017) A decision support system for automated identification of sleep stages from single-channel EEG signals. *Knowl Based Syst* 128:115–124. <https://doi.org/10.1016/j.knsys.2017.05.005>
- Hassan AR, Bashar SK, Bhuiyan MIH (2015) Automatic classification of sleep stages from single-channel electroencephalogram. In: 2015 annual IEEE India conference (INDICON)
- He G, Zhao W, Xia X et al (2019) An ensemble of shapelet-based classifiers on inter-class and intra-class imbalanced multivariate time series at the early stage. *Soft Comput* 23:6097–6114. <https://doi.org/10.1007/s00500-018-3261-3>
- Heyat MBB, Lai D, Khan FI, Zhang Y (2019) Sleep bruxism detection using decision tree method by the combination of C4–P4 and C4–A1 channels of scalp EEG. *IEEE Access* 7:102542–102553. <https://doi.org/10.1109/ACCESS.2019.2928020>
- Holland JV, Dement WC, Raynal DM (1974) Polysomnography: a response to a need for improved communication. In: Presented at the 14th annual meeting of the association of psychophysiological study of sleep
- Hsu YL, Yang Y-TC, Wang J-S, Hsu CY (2013) Automatic sleep stage recurrent neural classifier using energy features of EEG signals. *Neurocomputing* 104:105–114. <https://doi.org/10.1016/j.neucom.2012.11.003>
- Huang W, Guo B, Shen Y, Tang X, Zhang T, Li D, Jiang Z (2019) Sleep staging algorithm based on multichannel data adding and multifeature screening. *Comput Methods Programs Biomed*. <https://doi.org/10.1016/j.cmpb.2019.105253>
- Iber C (2007) *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*: American Academy of Sleep Medicine
- Kalbkhani H, Ghasemzadeh P, Shayesteh MG (2018) Sleep stages classification from EEG signal based on Stockwell transform. In: *IET Signal Processing* 2018. <https://ukdiiss.com/examples/sleep-stages-classification-eeeg-signal.php?vref=1>
- Khalighi S, Sousa T, Oliveira D, Pires G, Nunes U (2011) Efficient feature selection for sleep staging based on maximal overlap discrete wavelet transform and SVM. *Annu Int Conf IEEE Eng Med Biol Soc* 2011:3306–3309. <https://doi.org/10.1109/IEMBS.2011.6090897>
- Khalighi S, Sousa T, Pires G, Nunes U (2013) Automatic sleep staging: a computer assisted approach for optimal combination of features and polysomnographic channels. *Expert Syst Appl* 40(17):7046–7059. <https://doi.org/10.1016/j.eswa.2013.06.023>
- Khalighi S, Sousa T, Santos JM, Nunes U (2016) ISRUC-Sleep: a comprehensive public dataset for sleep researchers. *Comput Methods Programs Biomed* 124:180–192. <https://doi.org/10.1016/j.cmpb.2015.10.013>
- Kim T, Kim J, Lee K (2018) Detection of sleep disordered breathing severity using acoustic biomarker and machine learning techniques. *BioMed Eng OnLine* 17:16. <https://doi.org/10.1186/s12938-018-0448-x>
- Lai D, Heyat MBB, Khan FI, Zhang Y (2019) Prognosis of sleep bruxism using power spectral density approach applied on EEG signal of both EMG1–EMG2 and ECG1–ECG2 channels. *IEEE Access* 7:82553–82562. <https://doi.org/10.1109/ACCESS.2019.2924181>
- Memar P, Faradji F (2018) A novel multi-class EEG-based sleep stage classification system. *IEEE Trans Neural Syst Rehabil Eng* 26(1):84–95. <https://doi.org/10.1109/TNSRE.2017.2776149>
- Mendez MO, Chouvarda I, Alba A, Bianchi AM, Grassi A, Arce-Santana E, Milioli G, Terzano MG, Parrino L (2016) Analysis of A-phase transitions during the cyclic alternating pattern under normal sleep. *Med Biol Eng Comput* 54(1):133–148. <https://doi.org/10.1007/s11517-015-1349-9>
- Nagabushanam P, Thomas George S, Radha S (2019) EEG signal classification using LSTM and improved neural network algorithms. *Soft Comput*. <https://doi.org/10.1007/s00500-019-04515-0>
- Najdi S, Gharbali AA, Fonseca JM (2017) Feature transformation based on stacked sparse autoencoders for sleep stage classification. In: *Technological innovation for smart systems 2017*, pp 191–200. https://doi.org/10.1007/978-3-319-56077-9_18
- O'Reilly C, Nielsen T (2014) Assessing EEG sleep spindle propagation. Part 1: theory and proposed methodology. *J Neurosci Methods* 221:202–214. <https://doi.org/10.1016/j.jneumeth.2013.08.013>
- Obayya M, Abou-Chadi F (2014) Automatic classification of sleep stages using EEG records based on Fuzzy c-means (FCM) algorithm. In: *Radio science conference (NRSC), 2014 31st National*, pp 265–272. <https://doi.org/10.1109/NRSC.2014.6835085>
- Rahman T, Farook O, Heyat BB, Siddiqui MM (2016) An overview of narcolepsy. *IARJSET* 3:85–87
- Raza H, Cecotti H, Li Y et al (2016) Adaptive learning with covariate shift-detection for motor imagery-based brain-computer

- interface. *Soft Comput* 20:3085–3096. <https://doi.org/10.1007/s00500-015-1937-5>
- Rechtschaffen A (1968) A manual for standardized terminology techniques and scoring system for sleep stages in human subjects. *Brain Inf. Serv*
- Robnik-Šikonja M, Kononenko I (2003) Theoretical and empirical analysis of ReliefF and ReliefF. *Mach Learn* 53:23–69. <https://doi.org/10.1023/A:1025667309714>
- Ronzhina M, Janoušek O, Kolářová J, Nováková M, Honzík P, Provazník I (2012) Sleep scoring using artificial neural networks. *Sleep Med Rev* 16(3):251–263. <https://doi.org/10.1016/j.smrv.2011.06.003>
- Santaji S, Desai V (2020) Analysis of EEG signal to classify sleep stages using machine learning. *Sleep Vigil*. <https://doi.org/10.1007/s41782-020-00101-9>
- Schetinin V, Schult J (2006) Learning polynomial networks for classification of clinical electroencephalograms. *Soft Comput* 10:397–403. <https://doi.org/10.1007/s00500-005-0499-3>
- Şen B, Peker M, Çavuşoğlu A et al (2014) A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. *J Med Syst* 38:18. <https://doi.org/10.1007/s10916-014-0018-0>
- Sharma R, Pachori RB, Upadhyay A (2017) Automatic sleep stages classification based on iterative filtering of electroencephalogram signals. *Neural Comput Appl* 28:2959–2978. <https://doi.org/10.1007/s00521-017-2919-6>
- Sharma M, Tan R-S, Acharya UR (2019a) Automated heartbeat classification and detection of arrhythmia using optimal orthogonal wavelet filters. *Inf Med Unlocked*. <https://doi.org/10.1016/j.imu.2019.100221>
- Sharma M, Patel S, Choudhary S, Acharya UR (2019b) Automated detection of sleep stages using energy-localized orthogonal wavelet filter banks. *Arab J Sci Eng*. <https://doi.org/10.1007/s13369-019-04197-8>
- Siddiqui MM, Srivastava G, Saeed SH (2016) Diagnosis of insomnia sleep disorder using short time frequency analysis of PSD approach applied on EEG signal using channel ROC-LOC. *Sleep Sci* 9(3):186–191. <https://doi.org/10.1016/j.slsci.2016.07.002>
- Simões H, Pires G, Nunes U, Silva V (2010) Feature extraction and selection for automatic sleep staging using EEG. In: *Proceedings of the 7th international conference on informatics in control, automation and robotics*, vol 3, pp 128–133. <https://doi.org/10.5220/0002950601280133>
- Souri A, Ghafour MY, Ahmed AM et al (2020) A new machine learning-based healthcare monitoring model for student's condition diagnosis in Internet of Things environment. *Soft Comput* 24:17111–17121. <https://doi.org/10.1007/s00500-020-05003-6>
- Sousa T, Cruz A, Khalighi S, Pires G, Nunes U (2015) A two-step automatic sleep stage classification method with dubious range detection. *Comput Biol Med* 59:42–53. <https://doi.org/10.1016/j.compbimed.2015.01.017>
- Spriggs WH (2014) *Essentials of Polysomnography*. World Headquarter Jones & Bartlett Publishers, MA, United States
- Sriraam N, Padma Shri TK, Maheshwari U (2016) Recognition of wake-sleep stage 1 multichannel EEG patterns using spectral entropy features for drowsiness detection. *Australas Phys Eng Sci Med* 39(3):797–806. <https://doi.org/10.1007/s13246-016-0472-8>
- The Kappa statistic in reliability studies: use, interpretation, and sample size requirements (2005) *Physical Therapy*. <https://doi.org/10.1093/ptj/85.3.257>
- Tian P, Hu J, Qi J, Ye X, Che D, Ding Y, Peng Y (2017) A hierarchical classification method for automatic sleep scoring using multiscale entropy features and proportion information of sleep architecture. *Biocybern Biomed Eng* 37(2):263–271. <https://doi.org/10.1016/j.bbe.2017.01.005>
- Tzimourta KD, Tsimbaris A, Tzioukalia K, Tzallas AT, Tsiouras MG, Asktrakas LG, Giannakeas N (2018) EEG based automatic sleep stage classification. *Biomed J Sci Tech Res*. <https://doi.org/10.26717/BJSTR.2018.07.001535>
- Wang Q, Zhao D, Wang Y, Hou X (2019) Ensemble learning algorithm based on multi-parameters for sleep staging. *Med Biol Eng Comput*. <https://doi.org/10.1007/s11517-019-01978-z>
- Yasoda K, Ponnagall RS, Bhuvaneshwari KS et al (2020) Automatic detection and classification of EEG artifacts using fuzzy kernel SVM and wavelet ICA (WICA). *Soft Comput* 24:16011–16019. <https://doi.org/10.1007/s00500-020-04920-w>
- Zhang B, Lei T, Liu H, Cai H (2018) EEG-based automatic sleep staging using ontology and weighting feature analysis. *Comput Math Methods Med* 2018:1–16. <https://doi.org/10.1155/2018/6534041>
- Zhu G, Li Y, Wen P (2014) Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. *IEEE J Biomed Health Inform* 18(6):1813–1821. <https://doi.org/10.1109/JBHI.2014.2303991>
- Zibrantsen I, Kidmose P, Otto M, Ibsen J, Kjaer TW (2016) Case comparison of sleep features from ear-EEG and scalp-EEG. *Sleep Sci* 9(2):69–72

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.