Automated Classification of Sleep Stages Using Single-Channel EEG:
A Machine Learning-Based Method

Santosh Kumar Satapathy, Pondicherry Engineering College, India*

D. Loganathan, Pondicherry Engineering College, India

ABSTRACT

The main contribution of this paper is to present a novel approach for classifying the sleep stages based on optimal feature selection with ensemble learning stacking model using single-channel EEG signals. To find the suitable features from extracted feature vector, the authors obtained the ReliefF (ReF), Fisher Score (FS), and Online Stream Feature Selection (OSFS) selection algorithms. The research work was performed on two different subgroups of sleep data of ISRUC-Sleep dataset. The experimental results of the methodology signify that single-channel of EEG signal is superior to other machine learning classification models with overall accuracies of 97.93%, 97%, and 95.96% using ISRUC-Sleep subgroup-I (SG-I) data, and similarly, the model achieved overall accuracies of 98.16%, 98.78%, and 95.26% using ISRUC-Sleep subgroup-III (SG-III) data with FS, ReF, and OSFS, respectively.

KEYWORDS

Automated Sleep Scoring, Electroencephalography, Ensemble Stacking Model, Feature Screening, Machine Learning

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 (Panossian, 2009; Smaldone, 2007). Sleep disorders are common in all subsets of the population, independently of gender. This public health challenge greatly affects 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 some common examples of sleep-related disorders. 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 (Hassan, 2016). Sleep monitoring is one of the most significant activities in the assessment of 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, NREM sleep state is divided into four stages, namely NREM N1, N2, N3, and N4 (Aboalayon, 2014). 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 (Obayya, 2014).A sleep assessment can be supported by a sleep test with polysomnographic (PSG) recordings. PSG signals are a collection of different physiological signals that are collected from subjects during sleep.A PSG signal is combination of multivariate signal recordings, such as electroencephalogram (EEG), electrocardiogram (ECG), electrooculogram (EOG), and electromyogram (EMG) (Alickovic,2018) . The EEG signal recordings are used during sleep staging scoring. These signals represent brain activity, and, therefore, are suitable for evaluation of sleep abnormalities. After data collection, a sleep 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 (Abeyratne, 2007).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 (Rechtschaffen, 1968). 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) (Iber, 2007). The AASM manuals have combined the N3 and N4 stages into a single stage (N3) that is characterized by slow-wave sleep (SWS). Manual and visual evaluation of sleep scoring is complicated, costly and time-consuming. This manual approach overloads sleep experts who have to continuously monitor their patients during every sleep evaluation. However, this process provides the best accuracy in research on sleep disturbances (Bianchi, 2017). Consequently, the development of automated detection and recognition applications to assist sleep experts with the process of diagnosing sleep disorders is critical for enhanced public health. The EEG signals that are collected during sleep studies consist of either single-channel or multiple-channel recordings. EEG signals have several advantages over other types of signals. EEG signals can be obtained using wearable technologies that is comfortable for subjects. Moreover, the data collection process can be done in either the patient's home or a healthcare facility (Cogan, 2017).

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. More-over, these studies also suggest the use of different feature selection algorithms for selection of the most relevant features. Finally, different classification techniques have been used to analyze EEG signals considering two to six sleep stages. Multiple studies propose automatic sleep stage classification systems based on single-channel EEG signals. Oboyya et al. (Oboyya, 2014) proposed the use of single-channel EEG signals for sleep stage scoring in selected subjects between 35 and 50 years old. Moreover, they proposed wavelet transform techniques for feature extraction from EEG signals, and a fuzzy c-means algorithm for sleep stage classification. The reported overall classification accuracy was 85%. Güneş, K et al. (Güneş, 2010) designed an automatic sleep stage classification system was proposed based on a 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 et al. (Aboalayon, 2014) proposed 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 (Hassan, 2017) 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. Diykh, M et al. (Diykh, 2016) 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%. Gunnarsdottir, K. M et al. (Gunnarsdottir, 2018) designed an automated sleep stage scoring system using PSG data. The authors extracted both time and frequency domain properties from input channels such as EEG, EOG, and EMG. 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. (Sriraam, 2018) 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. (Memar, 2018) 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. In total, 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. (Da Silveira, 2016) proposed an automated sleep stage classification system based on single-channel EEG signals. The authors used discrete wavelet transform to decompose the signal into different sub-bands. 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%. Wutzl, B. et al. (Wutzl, 2019) proposed a hybrid automatic sleep stage classification system based on single-channel EEG signals. In this study, different extracted features from multiple domains such as time, frequency and non-linear features were forwarded to random forest classifiers. The overall classification accuracy of the proposed method was 85.95%. Zhu, G et al. (Zhu, 2014) 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. Braun, E. T. et al. (2018) proposed a portable sleep staging classification system using a different combination of features from EEG signals. The proposed method presented the best classification accuracy of 97.1% for the two-state sleep stage classification problem.

This research work proposes a new framework in the sleep staging classification problem which is quite different from the other contributions which are presented in the above. The proposed system generated the better performance in comparable to the other feature selection and ensemble learning methods. In this research work, an attempt is made to develop an efficient and robust automated sleep staging system using a single-channel EEG signal. The main objective of this study is to classify the sleep stages and improve the performance of sleep staging classification accuracy than other state-of-the-artwork. The main novelty of this work is obtaining ensemble learning with feature selection to improve the sleep staging classification accuracy. It has been seen from the previous studies that most of the researchers obtained the predefined features. But in this proposed work, we focused the feature selection with ensemble learning concept. Mainly the present research work is carried out into two major steps. In the first step, we identify the most relevant features from the extracted feature

vector by obtaining different feature selection algorithms. Secondly, the selected features were fed into the classifier ensemble learning stacking model for multi-class sleep stages classification problems by using gradient boosting and random forest (RF) algorithms. In the present study, we consider two different categories of subjects: 1) individuals affected with symptoms of mild sleep problems, from whom we obtained the sleep recordings 2) healthy control subjects without any sleep-related symptoms. The proposed classification model integrates features from both the classifiers. Finally, the efficiency of the proposed sleep staging system is compared with the other sleep staging classification studies currently contributed in the literature.

The rest of the paper is organized as follows: Section 2 describes the proposed methodology for sleep staging evaluation in detail. In Section 3, discusses the experimental results obtained from the proposed methodology from two subgroups of subjects. In Section 4, we briefly discuss the results of our proposed methodology, as well as its advantages and limitations, and compare the results with those of state-of-the-art methods. Section 5 ends with concluding remarks and a description of future work.

METHODOLOGY

In this research work, an effective and robust method is applied to classify the sleep stages automatically based on the selected optimal set of features with an ensemble learning stacking model. The main purpose of this study is to analyze the effectiveness of selected features with a combination of ensemble learning for multi-class sleep stages classification problems. The proposed approach considers two different categories of sleep recordings which include subjects' effects with different types of sleep-related disorders and the other category is subject having complete healthy control. The proposed work is divided into two phases, in the first phase identifying the suitable features from the extracted feature vector through obtaining different selection algorithms. In the second phase, an ensemble learning stacking model is considered for the classification of the sleep stages. The detailed layout of the main building blocks of the proposed work is shown in Fig. 1 and the detailed descriptions about this are explained below subsections.

EXPERIMENTAL DATA

To evaluate the proposed model we used single-channel EEG signals from one public datasets named as ISRUC-Sleep datasets.This dataset recorded and prepared under supervision of neurologists in the Hospital of Coimbra University (Khalighi,2016).The entire recordings were presented in the three different subgroups. In subgroup-I (SG-I), the recordings obtained from the subjects who were affected with different types of sleep-related disorders. In subgroup-III (SG-III), the recordings obtained from two different sessions in two different dates from the subjects having mild sleep problems. Finally in subgroup-III (SG-III), contained recordings of subjects having completely healthy controlled subjects.In this paper, we used only two subgroups (SG-I/SG-III) data. The proposed model evaluated from C3-A2 channel of EEG signal with a sampling rate of 200 Hz.The recorded sleep epochs are segmented into 30s epoch and the recorded epochs are labelled according to the AASM sleep scoring rules which includes Wake, N-1, N-2, N-3, and REM. Table 1 provides the detailed distribution of the EEG sleep stages.

Table 1. Distribution of sleep epochs

Subject Number/ Subgroups	W	N1	N2	N3	R
Subject-1(SG-I)	165	63	173	231	118
Subject-2 (SG-I)	231	72	226	147	74
Subject-9 (SG-I)	72	143	315	136	84
Subject-16 (SG-I)	128	125	280	120	97
Subject-1 (SG-III)	149	91	267	158	85
Subject-2 (SG-III)	89	120	274	149	118
Subject-5 (SG-III)	67	65	287	251	80
Subject-6 (SG-III)	54	111	261	247	77

Figure 1. The complete workflow of the proposed ensemble learning stacking model
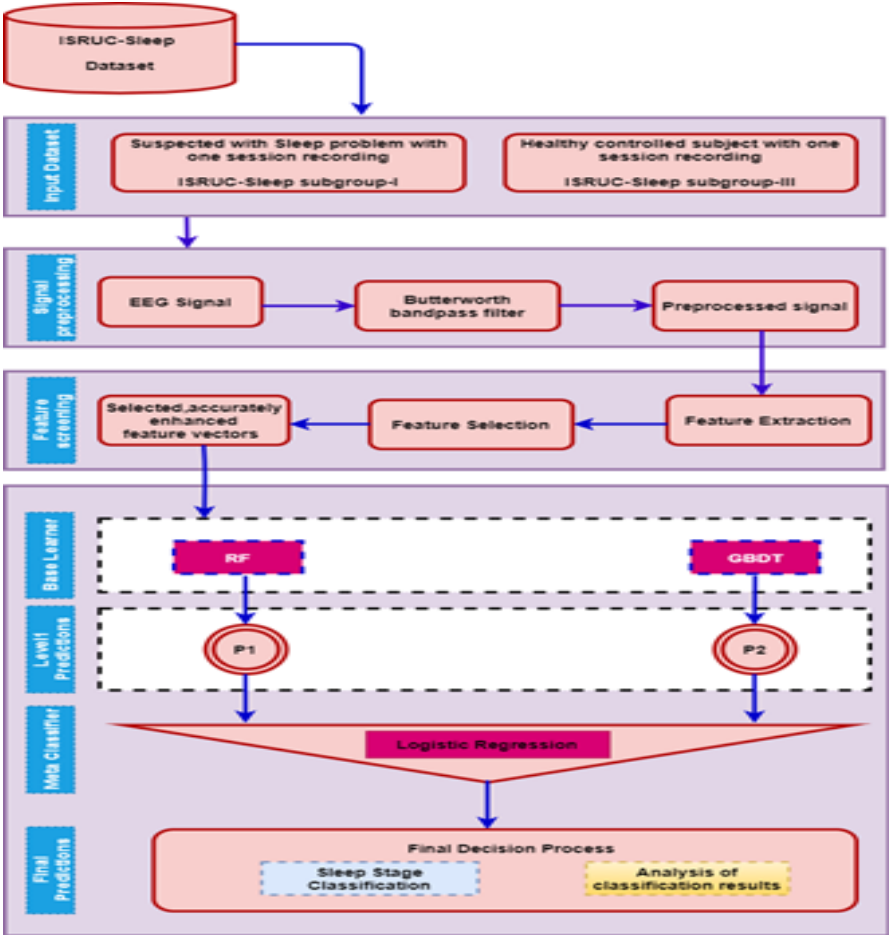
PRE-PROCESSING

In this research work, we segmented the EEG signal into epoch of 30s.The raw EEG signals are contaminated with the several types of irrelevant noise compositions and artefacts during recording hours. So it's highly important to eliminate these artefacts from the recorded signals for further processing. To reduce these artefacts, we used Butterworth band pass filter of order 10 with lower cutoff frequency 0.2 Hz and higher cutoff frequency 35 Hz.After that each epoch is normalized with zero mean and unit variance of the entire subject signal.

FEATURE EXTRACTION

It's quite difficult to directly analysis and identifying the irregularities from the pre-processed signal. For that reason, it's highly essential to retrieve the important information contained in the EEG signal into a reduced way of measurements or features. As we know that the sleep recordings are continuously changes its behavior with respect to time and frequency levels. So it's important to analysis the recordings with respect to time and frequency levels. In this study, we considered most of the already reported features. Mainly these features were classified into three categories such as time-domain, frequency-domain and time-frequency domain features. As a whole 30 features extracted from the signal which includes 12 time-domain features, 15 frequency-domain features, and 1 non-linear features.The extracted features described in Table 2.

Table 2. Extracted Features

Feature Domain	Extracted Feature	Feature Number
Time Domain	Mean, Maximum, Minimum, Standard Deviation Median, Variance, 75percentile, Signal Skewness, Signal Kurtosis	1-9
	Hjorth parameters	10-12
Frequency Domain	Relative Spectral Power δ, θ, α, β band	13-16
	Power Ratio δ/β, δ/θ, θ/α, θ/β, α/β, α/δ, $(\theta+\alpha)/(\alpha+\beta)$	17-23
	Band power in δ, θ, α, β	24-27
Non-linear	Zero-Crossing Rate	28

FEATURE REDUCTION

The main goal of this step is to select the most relevant features that help discriminate between the five sleep stages. It is sometimes the case that all extracted features are not suitable for a classification model, which may be one of the causes of degradation of the classification results. In this study, we used three different feature selection algorithms such as ReliefF (ReF) (Robnik-Šikonja, 2003), Fisher Score (FS) (Jin, 2016), and Online streaming feature selection (OSFS) (Eskandari, 2016) selection algorithm for screening the most suitable feature which helps to discriminate the changes sleep behavior over the individual sleep stages.

PROPOSED ENSEMBLE LEARNING STACKING CLASSIFICATION MODEL

In this research work, one of the primary contributions is an ensemble learning stacking model, which includes multiple classifiers systems that use a group of classifiers acts as base classifiers to build new training data to classify the sleep stages. First of all select, the base-layers classifiers say $B_1, B_2, \ldots\ldots, B_n$ train the classifiers using training dataset and generates multiple learners say $L_1, L_2, \ldots\ldots, L_n$ from training process. The output of these learners combined to prepare the new dataset in the form of $\left(y'0 \ldots\ldots\ldots y'm \right), yj$, which are fed into the next level classifier. Where $y'0$ is contained information of the predicted output of first base classifier of the input $x1'$ and yj is the actual output with input of $x1'$. The second layer of the model is called as meta-level classifier, which takes the output of the base-layer classifiers as an input and trained the model with help of this data. During the training process with the input set, the meta-layer identifies the errors of base-layer learners and adjusts them till the optimal solution is reached. This process continues for the k times to minimize the error and optimize the output performance. After continuity of this process, the meta-level classifier becomes tuned a generalization model which considers any type of input data. It has been seen that the traditional individual classifiers may perform poorly, that is the conventional classifier performed very well with training data but it has been noticed that the classifier nor performed well against the unseen new data. This challenge is overcome by introducing the ensemble learning stacking mode approach. The main benefit of this approach is, even though one of the classifiers from the base layers not performed well but averaging all the classifiers reduces the risk of depending on one individual classifier and it also avoids the poor selection of the classifier and helps to improve the classification accuracy. The complete pseudocode of the proposed ensemble learning stacking model is explained in Algorithm 1.

```
Algorithm 1
```
Input: `Training Dataset` $TD := \left(x_1, y_1 \right), \left(x_2, y_2 \right) \ldots\ldots\ldots\ldots \left(x_n, y_n \right)$
```
a
```
Output: `Final predictions results`
Step 1: `Learn base-layers classifiers (Level-1)`
```
Number of level-1 learners =2
```
Step 2: `Learn gradient boosting (TD, FE)`
```
        Assign NUTrees=50
```

$$F_0\left(a \right) = \arg\min \sum_{i=1}^{m} \left(\delta - y_i \right) = \arg\min \sum_{i=1}^{m} \left(\delta - y_i \right)^2$$

```
            Update the model based on number of target values
```
for `i=1 to` *m* **do**

$$F_0\left(a \right) = F_{i-1}\left(a \right) + G_{(i-1)}\left(a \right)$$

```
Calculate residue
``` $G_i\left( a \right)$
```
Where
``` $G_i\left( a \right) = predicted\ value - actual\ value$
```
end for
```
Step 3: `Learn Random Forest (DT, FE, NUTrees)`

$$G = \varnothing$$

```
for i=1 to NUTrees do
```
$$Tr^{(i)} = DT_i$$
```
for i=1 to
```
$Tr^{(i)}$
```
do
t=subset of FE
```
g_i =best feature in `t`

$$G \leftarrow G \cup g_i$$

```
end for
end for
```
Step 4: Construct new dataset of predictions to meta-level
classifier
```
for i=1 to n do
```

$$M_L = \left(x_1^{'}, y_1 \right)$$

Where $x_1^{'} = \left\{ L_1\left(x_1^{'}\right), \ldots\ldots\ldots\ldots L_n\left(x_1^{'}\right) \right\}$

```
end for
```
Step 5: Learn meta-level classifier (*E*)

Learn E based on M_L

Classify $x_1^{'}, x_2^{'}, x_3^{'} \ldots\ldots\ldots\ldots\ldots x_n^{'}$ as Wake, N1, N2, N3, REM stages

EXPERIMENTAL RESULTS

The proposed method is implemented using MATLAB R2015a software running on a personal laptop with an Intel Core™ i3-4005U CPU 1.70 GHz, 2 cores, 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 two different categories of subjects. The first experiment was conducted using data from subjects who were affected with different types of sleep problems and who had undergone a single sleep recording session. The second experiment was conducted using data from subjects who were completely healthy and no history of any diseases related to sleep. In this work we addressed the multiple classification tasks were involved in this study. These are the CT-2 (two-class), CT-3 (three-class), CT-4 (four-class) and CT-5 (five-class) discriminating problems. The effectiveness of the proposed methodology evaluated based on the different statistical parameters. In this work we have considered five parameters considered such as accuracy (Sanders, 2014), recall (Bajaj, 2013), specificity (Hsu, 2013), precision (Zibrandtsen, 2016), and F1score (Berry, 2014) for sleep staging performance evaluation. The extracted features were screened through three different feature selection techniques, FS, ReF and OSFS. The selected features are sorted according to feature weight (weights from high to low). Tables 3 and 4 presents the results of feature selection for the SG-I and SG-III data.

Table 3. Selected features: results with SG-I dataset

Selection Algorithm	Selected Feature Number (weights from high to low)
Fisher Score (FS)	18,8,13,10,5,15,17,4,1,14,9,11,7,2,19,23,16,24,22,26,12,25,3,21,20,6, 28,27
ReliefF Weight(ReF)	8,18,5,14,10,9,1,4,13,15,7,6,3,11,2,22,16,24,17,26,27,25,21,28,23,19,12,20
Online Stream Feature Selection (OSFS)	1,5,6,7,8,9,10,12,13,14,15,17,18,23,28,2,11,3,4,16,24,25,19,20,21,22,26

Table 4. Selected features: results with SG-III dataset

Selection Algorithm	Selected Feature Number(weights from high to low)
Fisher Score (FS)	1,17,16,24,22,2,15,14,3,10,28,26,6,11,21,20,27,19,23,25
ReliefF Weight(ReF)	13,18,8,12,1,5,7,4,9,17,15,2,10,14,3,16,24,22,6,28,11,23,19,20,27,21,25,26
Online Stream Feature Selection (OSFS)	1,3,4,5,8,9,13,14,15,16,17,18,22,2,10,6,12,7,11,19,20,25,23,24,21,28,27,26

PERFORMANCE OF SLEEP STAGING PERFORMANCE USING THE (SG-I) DATASET

In this experiment, we considered the single-channel C3-A2 of EEG signal recordings of four sleep-disordered subjects. All four subjects received a diagnosis of sleep syndrome disease, and some of the subjects also suffered from insomnia and other types of sleep problems. The selected features are forwarded to the two base layers classifiers that are random forest (RF) and gradient boosting decision tree (GBDT) to produce the initial sleep staging predictions. The classification accuracies (CAs) achieved for two to five sleep states classifications using the RF and GBDT classifiers are presented in Tables 5 and 6 respectively.

Table 5. Accuracy of the random forest classifier with the SG-I dataset

Feature Selection Algorithm	CT-2	CT-3	CT-4	CT-5
FS	99.22%	98.01%	97.86%	97.46%
ReF	99.01%	97.92%	97.79%	97.3%
OSFS	99.00%	97.97%	97.77%	96.96%

As shown in Table 5, the highest classification accuracy was reported for CT-2 (99.22%) with FS, CT-3 (98.01%) with FS, CT-4 (97.86%) with FS and CT-5 (97.46%) with FS selected features.

Table 6. Accuracy of the GBDT classifier with the SG-I dataset

Feature Selection Algorithm	CT-2	CT-3	CT-4	CT-5
FS	98.32%	97.91%	97.12%	95.2%
ReF	98.20%	97.79%	96.22%	94.86%
OSFS	98.01%	96.69%	95.70%	94.2%

As shown in Table 6, the highest classification accuracy was reported for CT-2 (98.32%) with FS, CT-3 (97.91%) with FS, CT-4 (97.12%) with FS and CT-5 (97.46%) with FS selected features.

SLEEP STAGING PERFORMANCE USING THE (SG-III) DATASET

This experiment looked at recordings from, four healthy control subjects who were not affected by any type of sleep-related problems. The same methodology was implemented for sleep staging with this dataset. The same properties were extracted to discriminate between the sleep classes. The performance achieved by the base learning classifiers using the SG-III dataset is presented in Tables 7 and 8 respectively.

Table 7. Accuracy of the RF classifier with the SG-III dataset

Feature Selection Algorithm	CT-2	CT-3	CT-4	CT-5
FS	99.52%	98.87%	97.36%	98.13%
ReF	99.11%	98.92%	96.79%	97.2%
OSFS	98.97%	97.89%	95.41%	96.56%

The highest accuracy was 99.52% with FS, 98.92% with ReF, 97.36% with FS, and 98.13% with FS for the CT-2 to CT-5 classification tasks.

Table 8. Accuracy of the GBDT classifier with the SG-III dataset

Feature Selection Algorithm	CT-2	CT-3	CT-4	CT-5
FS	98.72%	98.07%	97.56%	94.00%
ReF	98.21%	97.92%	96.69%	95.56%
OSFS	98.17%	97.49%	96.41%	96.3%

The highest classification accuracy was: CT-2, 98.72% with FS; CT-3, 98.07% with FS; CT-4, 97.56% with FS; and CT-5, 96.3% with OSFS.

Sleep staging performance using the proposed ensemble stacking algorithm with subgroup-I/III datasets

Finally, we deployed our proposed ensemble learning stacking model for sleep staging, which uses the base layer integration classification model. In this study, we used two base-layers classifiers, RF

and GBDT to make the first sleep scoring predictions. These first layer predictions were passed to the second layer, called the meta-classification layer. We applied the same feature selection technique combinations, which we finally obtained by performing the feature selection analysis. The classification results of this proposed ensemble learning stacking model using the SG-I and SG-III datasets for the two to five sleep state classification tasks are presented in Tables 9 and 10, respectively.

Table 9. Accuracy of the Ensemble Learning Stacking Model classifier with the SG-I dataset

Feature Selection Algorithm	CT-2	CT-3	CT-4	CT-5
FS	99.07%	98.02%	97.98%	97.93%
ReF	98.12%	97.42%	96.68%	97%
OSFS	98.97%	97.49%	96.61%	95.96%

The proposed ensemble learning stacking classification model showed the highest classification accuracy for CT-2 (99.07%), CT-3(98.02%), CT-4 (97.88%), and CT-5 (97.93%) using FS selected features.

Table 10. Performance evaluations with input from the SG-I dataset using the Ensemble Learning Stacking Model

Performance Metrics	SG-I Dataset (CT-5)			
	FS	RF	OSFS	
Accuracy	97.93%	97%	95.96%	
Precision	98.34%	96.66%	95.45%	
Sensitivity	97.71%	96.74%	95.42%	
F1-Score	98.03%	96.68%	95.43%	

As shown in Table 10, the highest classification accuracy of the proposed model was 97.93% with the FS selection algorithm. Similarly, the best performance in terms of precision (98.34%), recall (97.71%), and F1-score (98.03%) were reported with FS-selected features.

Table 11. Accuracy of the Ensemble Learning Stacking Model classifier with the SG-III dataset

Feature Selection Algorithm	CT-2	CT-3	CT-4	CT-5
FS	99.67%	99.02%	98.98%	98.16%
ReF	99.42%	99.22%	98.99%	98.78%
OSFS	99.17%	98.79%	97.61%	95.26%

The proposed stacking model achieved the highest classification accuracy of: CT-2, 99.67%, CT-3, 99.22%; CT-4, 98.99%; and CT-5, 98.78% with input from the SG-III dataset.

Table 12. Performance evaluation results with input from the SG-III dataset using the Ensemble Learning Stacking Model

Performance Metrics	SG-III Dataset (CT-5)			
	FS	**RF**	**OSFS**	
Accuracy	98.16%	98.78%	95.26%	
Precision	97.89%	96.97%	94.44%	
Sensitivity	97.94%	97.48%	94.89%	
F1-Score	97.91%	97.22%	94.66%	

As shown in Table 12, the proposed ensemble learning stacking model performed excellently on the five sleep state classification problem compared to the base layer classification algorithms for healthy control subjects. The proposed stacking model showed a classification accuracy of 98.78% using the ReF selection algorithm. Similarly, the highest precision (97.89%), sensitivity (97.94%), and F1-Score (97.91%) were achieved using the FS selection algorithm. Table 13 presents a summary of the results that were obtained through the different base layer classification models and the ensemble learning stacking model with the two different categories of subjects and the three feature selection techniques.

Table 13. Classification accuracy of various classification models and subject subgroups

Data Set	Classification Model	FS	ReF	OSFS
SG-I	**RF**	97.16%	97.3%	96.96%
	GBDT	95.2%	94.80%	94.2%
	Ensemble Learning Stacking Model	97.93%	97%	95.96%
SG-III	**RF**	98.13%	97.2%	96.56%
	GBDT	94.00%	95.56%	96.3%
	Ensemble Learning Stacking Model	98.16%	98.78%	95.26%

The best sleep staging was achieved using the ensemble learning stacking model with all three datasets. The highest classification accuracy was 99.34% for the five-sleep states classification problem with the SG-I dataset.

DISCUSSION

An ensemble learning stacking model based automated sleep scoring method is suggested for the classification of multiple sleep stages. In this research, we used only C3-A2 single-channel EEG signals to examine the subjects' sleep behaviour. According to the recent sleep staging studies, the C3-A2 channel is the most useful in terms of classification accuracy because it provides central information that represents the brain's behaviour during sleep (Hsu,2013; Zibrandtsen,2016;

Berry,2014; Liang,2012; Kim,2014;Peker,2016;Hassan,2017;Hassan,2017a;Diykh,2016;Diykh,201 6a;Mahvash Mohammadi,2016; Şen,2014; Obayya,2014; Radha,2014) . In other words, it conveys a good deal of information from the central part of the brain. Thus, we used C3-A2 channel recordings in the present work.The recorded signals are segmented into 30 s epochs with a sampling rate of 200 Hz. The proposed sleep staging experiment was executed according to AASM sleep scoring rules. The signals were pre-processed using a Butterworth bandpass filter. In the feature extraction step, we extracted both time and frequency domain features to characterize sleep behaviour in terms of frequency and time-oriented properties. The extracted feature vectors are forwarded to the ReF, FS, and OSFS feature selection techniques for further processing to select the suitable features for the classification models. In this work, we conducted two individual sleep staging experiments. The first two experiments were conducted on the sleep recordings from two different subgroups of subjects; these experiments used the base layer classification algorithms RF, and GBDT. The final experiment involved our proposed ensemble learning stacking model, where we used another layer for model learning, called the meta-classification layer. We compared the performance of the proposed system with that of other available state-of-the-art classification systems. To do so, we selected studies that used similar datasets and single-channel recordings. Table 14 compares the features used in the present work to others in related works, all of which relied on single-channel EEG signals from the ISRUC-Sleep dataset. These comparisons must take into account the use of single-channel EEG signals. Different features and classification models are presented in Table 15.

Table 14. Comparison of the CAs (%) of our proposed model with that of other state-of-the-art techniques using the same features and datasets

Author	Classifier	Classification Accuracy
Khalighi et al.,2011	SVM	95%
Hugo Simoes et al.,2010	Bayesian Classifier	83%
Khalighi et al.,2016	SVM	93.97%
Sousa et al.,2015	SVM	86.75%
Khalighi et al.,2013	SVM	81.74%
KD Tzimourta et al.,2018	Random Forest	75.29%
Najdi et al.,2017	Stacked Sparse Auto-Encoders (SSAE)	82.3%
Kalbkhani, H. et al.,2018	SVM	83.33%
Proposed	**Ensemble Learning Stacking Model (SG-I)**	**FS-97.93%**
		ReF-97%
		OSFS-95.96%
	Ensemble Learning Stacking Model (SG-III)	**FS-98.16%**
		ReF-98.78%
		OSFS-95.26%

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

Author	Classifier	Classification Accuracy
Huang, W. et al.,2019	SVM	92.04%
Dhok, S. et al.,2020	GSVM	87.45%
Wang, Q. et al.,2019	Stacking Model	96.6
Sharma, M. et al.,2019	SVM	91.5%
Hassan, A.R et al.,2016	Bagging	90.69%
Da Silveira et al.,2016	RF	91.5%
Memar, P. et al.,2018	RF	95.31%
Santaji, S. et al.,2020	RF	97.8%
Proposed	**Ensemble Learning Stacking Model (SG-I)**	**FS-97.93%**
		ReF-97%
		OSFS-95.96%
	Ensemble Learning Stacking Model (SG-III)	**FS-98.16%**
		ReF-98.78%
		OSFS-95.26%

CONCLUSION

This research work proposed an automated computer-aided sleep staging system for two to five sleep states classification problems based on ensemble learning stacking model using single-channel of EEG signal under the AASM sleep scoring guidelines. The proposed model was performed on two different subgroups data of ISRUC-Sleep dataset.The sleep recordings were acquired from the subjects who were affected with the various types of sleep-related disorders and healthy controlled subjects composed of 3000 epochs (SG-I), and 3000 epochs (SG-III) of length 30-s. This study makes three important contributions. First, extraction of numerous features from the time and, frequency domains as well as non-linear features. These sets of features support the analysis of sleep EEG parameters and their characteristics. Multi-feature extraction has improved the accuracy of sleep staging. Secondly, the proposed work used feature screening techniques, which are directly applicable to identifying the most relevant features from extracted feature vectors. In this work, we have applied the three different feature selection algorithms as the ReF, FS and OSFS to identify the most relevant features with proper screening, which makes more advantages for identifying the changes in sleep characteristics. Thirdly, the proposed work establishes an ensemble learning stacking model which integrates multiple classification models in two layers. The base layers consist of RF and GBDT classifiers, while the second layer contains a logistic regression algorithm. The proposed stacking model reported high recognition rates for multiple sleep staging classification tasks. The experimental results of the proposed methodology signify that single-channel of EEG signal superior to other machine learning classification models with overall accuracies of 97.93%, 97%, and 95.96% using SG-I data of ISRUC-Sleep dataset and similarly the proposed model achieved an overall accuracies of 98.16%, 98.78%, and 95.26% using SG-III data of ISRUC-Sleep dataset with FS, ReF and OSFS respectively. The results of this study show that the method provides an effective mechanism for handling data from subjects with different health conditions with high accuracy. We also compared the proposed research with similar studies from the literature and showed that the proposed model had better performance. On the other hand, the performances achieved in this study imply that the method can reliably aid sleep clinicians in the evaluation process and identification of sleep irregularities. Moreover, the proposed

system presents high performance for all categories of subjects with different medical conditions. We will extend our proposed work by integrating with an automated system based on a stacking ensemble approach based deep neural network model.

FUNDING AGENCY

REFERENCES

Abeyratne, U. R., Swarnkar, V., Rathnayake, S. I., & Hukins, C. (2007) Sleep-stage and event de-pendency of brain asynchrony as manifested through surface EEG. *Proceedings of the 29th IEEE Annual International Conference of the Engineering in Medicine and Biology Society*, 709–712.

Aboalayon, K. A. I., Ocbagabir, H. T., & Faezipour, M. (2014). Efficient sleep stage classification based on EEG signals. *IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014*.

Alickovic, E., & Subasi, A. (2018, June). Ensemble SVM method for automatic sleep stage classification. *IEEE Transactions on Instrumentation and Measurement*, *67*(6), 1258–1265.

Bajaj, V., & Pachori, R. (2013). Automatic classiðcation of sleep stages based on the time-frequency image of EEG signals. *Computer Methods and Programs in Biomedicine*, *112*(3), 320–328.

Berry, R. B., Brooks, R., Gamaldo, C. E., Hardsim, S. M., Lloyd, R. M., Marcus, C. L., & Vaughn, B. V. (2014). *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. American Academy of Sleep Medicine.

Bianchi, M. T. (2017). Sleep Devices: Wearables and nearables, informational and interventional. *Consumer and clinical. Metabolism: Clinical and Experimental*, *84*, 99–108.

Braun, E. T., Kozakevicius, A. D. J., Da Silveira, T. L. T., Rodrigues, C. R., & Baratto, G. (2018) Sleep Stages Classification Using Spectral Based Statistical Moments as Features. *Revista de Informática Teórica e Aplicada*, *25*(1).

Cogan, D., Birjandtalab, J., Nourani, M., Harvey, J., & Nagaraddi, V. (2017). Multi-Biosignal Analysis for Epileptic Seizure Monitoring. *International Journal of Neural Systems*, *27*(1), 1650031.

Da Silveira, T. L. T., Kozakevicius, A. J., & Rodrigues, C. R. (2016). Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain. *Medical & Biological Engineering & Computing*, *55*(2), 343–352.

Dhok, S., Pimpalkhute, V., Chandurkar, A., Bhurane, A. A., Sharma, M., & Achary, U. R. (2020). Automated phase classification in cyclic alternating patterns in sleep stages using Wigner-Ville Distribution based features. *Computers in Biology and Medicine*, *119*, 103691.

Diykh, M., & Li, Y. (2016). Complex networks approach for EEG signal sleep stages classification. *Expert Systems with Applications*, *63*, 241–248.

Diykh, M., Li, Y., & Wen, P. (2016). EEG Sleep Stages Classification Based on Time Domain Features and Structural Graph Similarity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *24*(11), 1159–1168.

Diykh, M., Li, Y., & Wen, P. (2016a). EEG sleep stages classification based on time domain features and structural graph similarity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *24*(11), 1159–1168.

Eskandari, S., & Javidi, M. M. (2016). Online streaming feature selection using rough sets. *International Journal of Approximate Reasoning*, *69*, 35–57.

Güneş, S., Polat, K., & Yosunkaya, Ş. (2010). Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting. *Expert Systems with Applications*, *37*(12), 7922–7928.

Gunnarsdottir, K. M., Gamaldo, C. E., Salas, R. M. E., Ewen, J. B., Allen, R. P., & Sarma, S. V. (2018). 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)*.

Hassan, A. R., & Bhuiyan, M. I. H. (2016). Computer-aided sleep staging using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and bootstrap aggregating. *Biomedical Signal Processing and Control*, *24*, 1–10.

Hassan, A. R., & Bhuiyan, M. I. H. (2017). An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting. *Neurocomputing*, *219*, 76–87.

Hassan, A. R., & Bhuiyan, M. I. H. (2017a). Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting. *Computer Methods and Programs in Biomedicine*, *140*, 201–210.

Hassan, A. R., & Hassan Bhuiyan, M. I. (2016). Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. *Biocybernetics and Biomedical Engineering*, *36*(1), 248–255. doi:10.1016/j.bbe.2015.11.001

Hassan, A. R., & Subasi, A. (2017). A decision support system for automated identification of sleep stages from single-channel EEG signals. *Knowledge-Based Systems*, *128*, 115–124.

Hsu, Y.-L., Yang, Y.-T., Wang, J.-S., & Hsu, C.-Y. (2013). Automatic sleep stage recurrent neural classiðer using energy features of EEG signals. *Neurocomputing*, *104*, 105–114.

Huang, W., Guo, B., Shen, Y., Tang, X., Zhang, T., Li, D., & Jiang, Z. (2020, April). Sleep staging algorithm based on multichannel data adding and multifeature screening. *Computer Methods and Programs in Biomedicine*, *187*, 105253.

Iber, C., Ancoli-Israel, S., Chesson, A. L., & Quan, S. F. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*. American Academy of Sleep Medicine.

Jin, X., Bo, T., He, H., & Hong, M. (2016). Semi supervised feature selection based on relevance and redundancy criteria. *IEEE Transactions on Neural Networks and Learning Systems*, *28*, 1974–1984.

Kalbkhani, H., Ghasemzadeh, P. G., & Shayesteh, M. (2018). Sleep Stages Classification from EEG Signal based on Stockwell Transform. *IET Signal Processing*.

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. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3306-3309.

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 Systems with Applications*, *40*(17), 7046–7059.

Khalighi, S., Sousa, T., Santos, J. M., & Nunes, U. (2016). ISRUC-Sleep: A comprehensive public dataset for sleep researchers. *Computer Methods and Programs in Biomedicine*, *124*, 180–192.

Khalighi, S., Sousa, T., Santos, J. M., & Nunes, U. (2016). ISRUC-Sleep: A comprehensive public dataset for sleep researchers. *Computer Methods and Programs in Biomedicine*, *124*, 180–192.

Kim, J. (2014). A Comparative Study on Classification Methods of Sleep Stages by Using EEG. *J. Korea Multimed. Soc.*, *17*(2), 113–123.

Liang, S.-F., Kuo, C.-E., Hu, Y.-H., Pan, Y.-H., & Wang, Y.-H. (2012). Automatic Stage Scoring of Single-Channel Sleep EEG by Using Multiscale Entropy and Autoregressive Models. *IEEE Transactions on Instrumentation and Measurement*, *61*, 1649–1657.

Mahvash Mohammadi, S., Kouchaki, S., Ghavami, M., & Sanei, S. (2016). Improving time–frequency domain sleep EEG classification via singular spectrum analysis. *Journal of Neuroscience Methods*, *273*, 96–106.

Memar, P., & Faradji, F. (2018). A Novel Multi-Class EEG-Based Sleep Stage Classification System. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(1), 84–95.

Najdi, S., Gharbali, A. A., & Fonseca, J. M. (2017). Feature Transformation Based on Stacked Sparse Auto encoders for Sleep Stage Classification. *Technological Innovation for Smart Systems*, 191–200.

Obayya, M., & Abou-Chadi, F. E. Z. (2014). Automatic classification of sleep stages using EEG records based on Fuzzy c-means (FCM) algorithm. *2014 31st National Radio Science Conference (NRSC)*, 265–272.

Obayya, M., & AbouChadi, F. (2014) Automatic classification of sleep stages using EEG records based on Fuzzy c-means (FCM) algorithm. *Radio Science Conference (NRSC)*, 265-272.

Obayya, M., & Abou-Chadi, F. (2014). Automatic classification of sleep stages using EEG records based on Fuzzy c-means (FCM) algorithm. *National Radio Science Conference, NRSC, Proceedings*, 265-272.

Panossian, L. A., & Avidan, A. Y. (2009). Review of Sleep Disorders. *The Medical Clinics of North America*, *93*(2), 407–425. doi:10.1016/j.mcna.2008.09.001 PMID:19272516

Peker, M. (2016). A new approach for automatic sleep scoring: Combining Taguchi based complex-valued neural network and complex wavelet transform. *Computer Methods and Programs in Biomedicine*, *129*, 203–216.

Radha, M., Garcia-Molina, G., Poel, M., & Tononi, G. (2014). Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal. *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014*, 1876–1880.

Rechtschaffen, A., & Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring Systems for Sleep Stages of Human Subjects*. U.G.P. Office, Public Health Service.

Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and Empirical Analysis of ReliefF and ReliefF. *Machine Learning*, *53*, 23–69.

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

Santaji, S., & Desai, V. (2020). *Analysis of EEG Signal to Classify Sleep Stages Using Machine Learning*. Sleep and Vigilance.

Şen, B., Peker, M., Çavuşoğlu, A., & Çelebi, F. V. (2014). A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms. *Journal of Medical Systems*, *38*(3), 18.

Shabani, F., Kumar, L., & Solhjouy-Fard, S. (2017). Variances in the projections, resulting from CLIMEX, boosted regression trees and random forests techniques. *Theoretical and Applied Climatology*, 1–14.

Sharma, M., Patel, S., Choudhary, S., & Acharya, U. R. (2019). Automated Detection of Sleep Stages Using Energy-Localized Orthogonal Wavelet Filter Banks. *Arabian Journal for Science and Engineering*.

Simões, H., Pires, G., Nunes, U., & Silva, V. (2010) Feature Extraction and Selection for Automatic Sleep Staging using EEG. *Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics*, *3*, 128-133.

Smaldone, A., Honig, J. C., & Byrne, M. W. (2007). Sleepless in America: Inadequate sleep and relationships to health and well-being of our nation's children. *Pediatrics*, *2007*(119), 29–37. doi:10.1542/peds.2006-2089F PMID:17272582

Sousa, T., Cruz, A., Khalighi, S., Pires, G., & Nunes, U. (2015). A two-step automatic sleep stage classification method with dubious range detection. *Computers in Biology and Medicine*, *59*, 42–53.

Sriraam, N., Padma Shri, T. K., & Maheshwari, U. (2018). Recognition of wake-sleep stage 1 multichannel EEG patterns using spectral entropy features for drowsiness detection. *Australasian Physical & Engineering Sciences in Medicine*, *39*(3), 797–806.

Tang, B., Chen, Q., & Wang, X. (2010) Reranking for stacking ensemble learning. *International Conference on Neural Information Processing: Theory and Algorithms*, 575–584

Tzimourta, K. D., Tsilimbaris, A. K., Tzioukalia, A. T., Tzallas, M. G., & Tsipouras, L. G. (2018) EEG-Based Automatic Sleep Stage Classification. *Biomed J Sci &Tech Res*, *7*(4).

Wang, Q., Zhao, D., Wang, Y., & Hou, X. (2019). Ensemble learning algorithm based on multi-parameters for sleep staging. *Medical & Biological Engineering & Computing*, *57*(8), 1693–1707.

Wutzl, B., Leibnitz, K., Rattay, F., Kronbichler, M., Murata, M., & Golaszewski, S. M. (2019). Genetic algorithms for feature selection when classifying severe chronic disorders of consciousness. *PLoS One*, *14*(7), e0219683.

Xi, J., & Coggeshall, S. (2010). Prediction of transfers to tertiary care and hospital mortality: A gradient boosting decision tree approach. *Statistical Analysis and Data Mining*, *3*, 253–258.

Zhu, G., Li, Y., & Wen, P. P. (2014). Analysis and Classification of Sleep Stages Based on Difference Visibility Graphs From a Single-Channel EEG Signal. *IEEE Journal of Biomedical and Health Informatics*, *18*(6), 1813–1821.

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

Santosh Kumar Satapathy was born in Parlakhemundi, India, in 1987. He received the B.Tech. Degree in Information Technology from the BPUT, Odisha, India, in 2009, and the M.Tech. Degree in Computer Science and Engineering from the BPUT, Odisha, India, in 2012.In 2012, he joined the Department of Information Technology, Roland Institute of Technology, Berhampur, Odisha, as a Lecturer, and in 2013 he joined as an assistant professor in the Department of Computer Science and Engineering, Gandhi Engineering College, Bhubaneswar, Odisha. He is currently a PhD student in the Department of Computer Science and Engineering at Pondicherry Engineering College, Puducherry. His research interests include data mining, biomedical signal analysis, and machine learning.

D. Loganathan is a Professor in the Department of CSE, Pondicherry Engineering College, Puducherry, India. He has over 32 years of teaching experience. He obtained his PhD from Information & Communication Engineering Faculty, College of Engineering, Anna University, Chennai, INDIA in 2008.He obtained Graduated and Post-Graduated with specialization in Computer Science from Birla Institute of Technology and Science, Pilani, Rajasthan, India in an integrated way from 1978 – 1984.He has published over 30 technical papers in different international/ national journals/conference proceedings and is serving in the editorial board of several international/national journals. He has contributed two chapters in a leading scientific and technical book. His current research interest includes Image Processing and Information Security.Dr. Loganathan is a Life Member of the Indian Society for Technical Education (ISTE), Computer Society of India (CSI), Institute of Engineers (India) and Indian Society for Remote Sensing. He has deputed for TEN week training at JAPAN in the CICC Tokyo, Japan in Jan – Mar, 1994. He has acted as Panel Member for AICTE and NBA, of Govt. of India, for a number of times in evaluating new and established Educational institutes.