Performance analysis of machine learning algorithms on automated sleep staging feature sets

Santosh Satapathy | D Loganathan | Hari Kishan Kondaveeti | RamaKrushna Rath

Abstract
With the speeding up of social activities, rapid changes in lifestyles, and an increase in the pressure in professional fields, people are suffering from several types of sleep-related disorders. It is a very tedious task for clinicians to monitor the entire sleep durations of the subjects and analyse the sleep staging in traditional and manual laboratory environmental methods. For the purpose of accurate diagnosis of different sleep disorders, we have considered the automated analysis of sleep epochs, which were collected from the subjects during sleep time. The complete process of an automated approach of sleep stages’ classification is majorly executed through four steps: pre-processing the raw signals, feature extraction, feature selection, and classification. In this study, we have extracted 12 statistical properties from input signals. The proposed models are tested in three different combinations of features sets. In the first experiment, the feature set contained all the 12 features. The second and third experiments were conducted with the nine and five best features. The patient records come from the ISRUC-Sleep database. The highest classification accuracy was achieved for sleep staging through combinations with the five feature set. From the categories of the subjects, the reported accuracy results were found to exceed above 90%. As per the outcome from the proposed system the random forest classification techniques achieved best accuracy incomparable to that of the other two classifiers.

1 INTRODUCTION
The analysis and study of human brain behaviour can be very important in different health sector applications such as in diagnosis of mental and neuro-disorder diseases and abnormalities. The activities of the human brain have been analysed through recordings of the EEG signal. The major role of EEG is to assist the physicians for accurate diagnosis of various diseases from recorded behaviour of the subjects which assist the sleep experts in distinguishing the sleep irregularities and prediction of epileptic seizures etc. [1] Besides this, sleep-related disorders are currently the major global challenge across the globe because a proper sleep can decide the major balance between our physical and mental health, as it directly affects the human quality of life and its bad consequences might reflect in our body [2]. Therefore the sleep study gives more insights into various health risks and issues such as memory impairment, diabetes, and cardiovascular diseases [3]. Henceforth, it is more vital for analysis of possible symptoms of sleep problems, and a sleep staging score is necessary for almost all conditions.

The procedure of sleep scoring is to identify the irregularities that happen with the help of polysomnography recordings from the patients during sleep hours. Generally, polysomnography is one of the multi-parameter tests and it helps in the analysis and interpretation of multiple and simultaneous activities happening in the body during sleep through recordings of different electrophysiological signals. During the treatment, different physiological signals have been recorded from different parts of the body. The brain behaviour information is collected through recordings of the electroencephalogram (EEG) information; similarly the chin, left and...
right limb activities are recorded using the signal from electromyogram (EMG), the behaviour of eye movements is acquired through the electrooculogram (EOG) signal recording, and the heart rhythm information is collected using the electrocardiogram (ECG) signal along with respiratory signals, airflow signals, and oxygen saturation [4]. The sleep experts measured all these recordings and analyses through standard sleep manuals named as R&K rules, which were documented by two scientists namely Rechtschaffen and Kales [5]. As per the R&K rules, whole sleep is classified into three stages—awake(W), non-rapid eye movement (NREM) and rapid eye movement (REM). Furthermore, NREM sleep stages are divided into N1-N4 sub sleep stages. N1 and N2 were considered as light sleep, similarly N3 and N4 as deep sleep. According to the R&K rules, each epoch length is the 30-s data. Furthermore, another sleep standard was published in the year 2007 named as AASM with minor revision to the earlier R&K guidelines [6]. According to AASM, the whole sleep was segmented into five sleep stages. The only changes observed between AASM and the R&K rule are the total number of sleep stages. In AASM, the N-REM stage3 and N-REM stage4 are combined into one stage called the N-REM stage3; another sleep stage was the same as that in the R&K rules. From all the electrophysiological signals, only EEG signals are used mostly subjected to the treatment and diagnosis of possible sleep-related diseases because they directly provide the information on brain activities during sleep. Another major advantage of EEG signals is that they are characterized by individual sleep stages through different EEG waveforms such as alpha, theta, beta, delta, and gamma [7]. The brain-behaviour is recorded from the scalp by attaching the electrodes, which mainly track the changes in voltage differences between the different neurons. The placement of electrodes is managed according to the international 10–20 EEG electrode placement rules. The concept of 10–20 refers to the placement of the electrodes which were either 10% and 20% right-left or front-back of the scalp [8]. Nowadays for the diagnosis of different types of sleep disorder, EEG signal is preferable in the sleep laboratory for monitoring the sleep quality in the real-time environment and sometimes it may also used in a portable manner at home so that one can easily measure the sleep quality [9].

The wake stage is considered a relaxed state where one subject can prepare himself or herself for sleeping. After sometime the body enters the N-REM stage1 and from this step, a sleep cycle begins; it is a transition phase between being awake and asleep. The average duration of this state is 5 to 15 min in general and the brain normally produces theta waves [10]. In N-REM stage2 is somewhat a deeper sleep compared to stage1, where the heart rhythm and body temperature reduce and sleep spindles and k-complexes also occur [11]. In the N-REM stage3, the actual deep sleep starts, and the brain behaves very slow and similar conditions continue for the N-REM stage4, wherein delta waves are produced from the brain. Finally, in the REM stage, small brain waves are seen and the blood pressure may suddenly increase with irregular breathing with rapid eye movements occurring in this phase of sleep. This process continues cyclically throughout the whole night from NREM stages to the REM stage and each cycle duration approximately lasts for 90 min [12].

One healthy sleep covers normally three to five sleep cycles for subjects. In general, the recordings of the EEG signals are highly in complicated in nature and the characteristics of these signals continuously fluctuate with regard to the amplitude, phase, and frequency. To extract meaningful data from raw EEG signals, we need to analyse the extended signal periods [13]. Another aspect of difficulty also observed in terms of EEG recording is that generally recorded hours of sleep data is quite difficult to analyse within the 5–10 s time window [14].

Sometimes, it has observed from different sleep studies that the multi-channel EEG approach raised some limitations towards the evaluation of irregularities due to disturbances found on the subject's health. Therefore most of the sleep researchers used a single-channel of EEG for classifying the sleep stages.

Traditionally, the technicians have analysed and interpreted the sleep stages manually, alternatively, it has raised the set of limitations which are related to manage the huge amount of sleep EEG records and it is also a time-consuming, highly expensive process and finally, this manual interpretation of sleep records is a human-dependent process [15]. As an outcome, it has required to develop an automated sleep stage classification system to get better classification accuracy.

In the present research on sleep staging, we propose an automated sleep staging system based on single-channel EEG signals from the subject with different health conditions. The proposed research completed majorly in five stages such as signal preprocessing, feature extraction, feature screening, classification algorithm, and finally the comparative analysis in between was from different categories of subjects.

In our proposed sleep study, we have considered three different category of feature combinations; in one category we have considered 12 features, the second category, we obtained the best nine features from the extracted feature vector and finally, in the third category we obtained the best five features. The novelty of this research work is the analysis of the sleep behaviour of the subjects through the statistically significant difference among the sleep states.

The rest of the paper is organized as follows: Section 2 describes the detailed on the existing contributions. The experimental data preparation is explained in Section 3. Section 4 described full description on the methodologies including data preprocessing, feature extraction, and feature selection, classification and model performance evaluation. Section 5 discusses the obtained experimental results of the proposed methodology from three different combinations of feature sets. Section 6 ends with the concluding remarks and the future work description.
2 | RELATED WORK

Recently, different sleep studies have been proposed by different authors which were subjected to the diagnosis of sleep-related disorders. Previously contributed different sleep staging was based on the EEG recordings and different signal analysis techniques. It has found that the automatic sleep stage classification is a challenging task with subject to proper managing the EEG signal processing in different levels, segmentation of the sleep epochs, extracting the features, screening the features and finally the selection of suitable classification models. In majority of our literature survey, we have focused on the consideration of the EEG signal, different types of features extracted and classification models obtained by different authors during the sleep irregularities analysis. Here, we described some of the existing related works with regard to sleep scoring in which different methodologies were obtained by different researchers.

Statistical features such as arithmetic mean [1], maximum and minimum values [13], standard deviation [2], median [3], skewness [4], kurtosis [5], variance [6], zero crossing value [7], Hjorth activity, mobility, complexity [8], mean energy [9], approximate entropy [10], spectral entropy [16], multiscale entropy [17], permutation entropy [18], renyi entropy [19], wavelet transform [20], band power (delta to beta range) [17,21–24], and mean frequency for different frequency bands [25–27], were used.

Flexer et al. obtained a hidden Markov model for classifying three stages of sleep and achieved an overall classification accuracy of 80% [28].

Berthier et al. proposed two to five sleep stage classifications with the input of EEG signals through the fuzzy-logic iterative methods and the accuracy was reported as 82.9% [29].

Chapotot and Becq [30] proposed the classification of six different sleep stages (wake, S1, S2, S3, S4, and REM) using a hybrid model in the combination of decision rules and k-nearest neighbor (KNN). The model achieved an overall accuracy of 78%.

Sinha presented the sleep study with consideration of the three-state sleep stage scoring and obtained wavelet transform techniques for signal segmentation and used ANN techniques for classification and reported the classification accuracy as 95.55% [31].

Ebrahimi et al. obtained the wavelet concepts for discriminating the different sleep stages and also used an ANN classifier reported accuracy rate as 93% [32].

Ronzhina et al. conducted the sleep study with consideration of power spectral density features and used the ANN classifier for six-state sleep stage classification and the accuracy was reported as 82.9% [33].

Jo et al. proposed a four-state sleep stage classification and obtained a genetic-fuzzy classifier for classification and finally, the model achieved 84.6% overall classification accuracy [34].

Hsu et al. proposed a sleep stage classification with the input of sleep EEG record and extracted energy characteristics for distinguishing the sleep stages. The accuracy result reported for the five-state sleep stage classification was 87.2% using a neural network model [35].

Fraiwan et al. extracted time-frequency entropy features to represent the sleep records and used a linear discriminate analysis algorithm for classifying the sleep stages and the achieved an accuracy result with six-state sleep stages as 84% [11].

Eduardo T. Braun presented an efficient and effective sleep staging scoring system and obtained frequency domain properties. They obtained random forest (RF) techniques and the achieved results for a classification model as 90.9%, 91.8%, 92.4%, 94.3%, and 97.1% for 2–6 sleep stages respectively [36].

In [37] the author has considered the graph theory concept for distinguishing the sleep stages by implementing the concepts of difference visibility graph and visibility graph from a single channel of EEG signal. The sleep-EDF dataset used for this research work obtained the support vector machine classifier for classifying the 2–6 sleep stage classification and the model achieved an average accuracy of 87.5%.

S.-F. Liang has considered sleep recordings from the public Sleep-EDF database and used linear discriminate analysis classifier for classification and achieved the accuracy result as 83.6% [2].

In [38] the sleep study has performed with EEG signals of the Sleep-EDF dataset, and the input signal was segmented into five sub-bands such as delta, theta, alpha, beta, and gamma through butter-worth band-pass filter techniques. Energy, standard deviation, and entropy features were extracted from the respective five sub-bands. Here, the author obtained an SVM classifier for the two-state sleep stage classification and achieved a success rate of 92%.

Hassan et al. presented six different studies related to sleep stage scoring using the EEG signal. In their first work [39], he extracted statistical moment features from a single channel of EEG signals and used a decision tree (DT) classification algorithm for classification. The reported accuracy result for the respective 6–2 sleep stages is 86.8%, 90.6%, 92.1%, 94.1%, and 99.4%.

In [40] the author presented another novel method towards the classification of sleep stages by introducing the empirical mode decomposition for decomposing the EEG signal into different segments. The extracted set of statistical features is forwarded into an adaptive boosting classifier. The model reported an overall accuracy for 6–2 sleep stages is 88.62%, 90.11%, 91.2%, 93.55%, and 97.73%.

In [41] the author has conducted sleep studies by introducing the empirical mode decomposition for signal decomposition into different frequency sub-bands. The whole experiments were conducted with Sleep-EDF sleep EEG records and selected features are forwarded into a Bagged classification tree and the classification model achieved an overall accuracy for 6–2 state sleep stages that are indicated as 86.89%, 90.69%, 92.14%, 94.10%, and 99.48%.

In [42] the overall accuracy reached was 90.3%, 91.5%, 92.1%, 94.8% and 97.5% with the input channel Pz-Oz; similarly for Fpz-Cz the result reported as 87.1%, 88.5%, 90.6%, 92.3% and 96.5% for 6-2 stages respectively.

In [43], the author has implemented the same techniques as TQWT for signal segmentation and extracted normal inverse
Gaussian features from respective signal sub-bands and used AdaBoost classification techniques for classification and the classification accuracy performance reported was 90.01%, 91.36%, 92.46%, 94.83% and 98.01% for 6–2 sleep state classifications.

In [44] the author used two bench-mark sleep dataset named as Sleep-EDF and Dream dataset, same TQWT techniques used for signal segmentation and bagging classification techniques and the accuracy result reported for 2-6 sleep state classification as 92.43%, 93.69%, 94.36%, 96.55%, and 99.75%.

Subasi et al. has proposed three sleep stages (alert, drowsy, sleep) using the EEG signal and obtained a wavelet transform concept for signal decomposition. The high classification accuracy achieved with the ANN classification algorithm and the classification model reported accuracy for alert, drowsy, and sleep were 92.3%, 96.2%, and 93.6% [2].

3 | EXPERIMENTAL DATA DESCRIPTION

3.1 | Experimental data

In this research work, we have used sleep recordings that were obtained from the ISRUC-Sleep public repository, which was directly prepared by sleep experts at Hospitalar and University of Coimbra [14]. This dataset was exclusively prepared for sleep-related analysis. In this dataset, three different subgroups of data were contained which was obtained from different medical conditioned subjects, from subgroup-I section, total 100 subjects one session recordings details were contained, similarly in a subgroup-II section, data of eight subjects were contained with two recording session per subjects and finally in a subgroup-III, data of 10 healthy subjects were collected as one session recording from each subject. All these recordings were recorded from subjects through a whole-night that include six EEG channels, two EOG channels and three EMG channels. The acquired signal recordings were performed with a sampling rate of 200 Hz and these recordings were obtained through a 10-20 international standard electrode placement system. In the present study, the major focus to analyse the sleep irregularities was through proper sleep stage classifications with the input of a single-channel consisting of C3-A2 of EEG signals. The C3-A2 channel was selected in most recent sleep studies and several studies [2, 41, 45–60] were achieved higher classification accuracies with the input of the C3-A2 channel.

In this study, we have used two different subgroups data from the ISRUC-sleep dataset, where four subjects are from the ISRUC-sleep subgroup-I, considering that the subjects were affected with different sleep problems, whereas other four data obtained from the ISRUC-sleep subgroup-III section, where the subjects were completely healthy controls. In this work, we have obtained eight polysomnography (4 subjects 1PSG + 4 subjects. 1PSG = 8 PSGs) records. Tables 1 and 2 presented the different sleep stages epochs from sleep disordered subject and subjects were completely healthy controls.

3.2 | Manual methods of sleep scoring

Each polysomnography recordings is associated with the corresponding annotation file, which are prepared according to AASM rules through two well-trained physicians which were labelled as the epochs in any of the sleep stage. Each epoch of the recordings is segmented into 30 s epochs.

3.3 | Data preparation

The recorded overnight PSG recordings were divided into 30-s long epochs. From each subject, a rectangular length of 6000 samples (=30 s × 200 Hz) of EEG signal segments are extracted. For each subject we prepared individual datasets $D = \{D1, D2, D3, D4, D5, D6, D7, D8\}$ with dimensions $i \times j$ where $i$ refers to total epochs and $j$ points to the length of the sample points. Let $D_{mn}$, $i = 1,2,3…….,m$, $j = 1,2,3…….,n$ be the element of the data matrix $D$. In this step, the length of $j$ is 6000 samples $30 \, s \times Fs = 6000$.

4 | PROPOSED METHODOLOGY

The complete research work layout of this proposed study is presented in Figure 1. It can be observed that the whole experiment is implemented through different phases. The first phase is carried out with proper signal preprocessing; the second phase with subject for analysis of characteristics of features, the third phase by selecting the suitable features through the feature selection algorithm; and finally the last

| TABLE 1 | Experimental information on the sleep disorder affected subject |
|----------------|------------------|
| Sleep stages   | Number of segments |
| Wake           | 847              |
| N1 stage       | 446              |
| N2 stage       | 881              |
| N3 stage       | 434              |
| REM            | 392              |

Abbreviation: REM, rapid eye movement.

| TABLE 2 | Experimental information on healthy control subjects |
|----------------|------------------|
| Sleep Stages   | Number of Segments |
| Wake           | 648              |
| N1 stage       | 349              |
| N2 stage       | 942              |
| N3 stage       | 674              |
| REM            | 387              |

Abbreviation: REM, rapid eye movement.
phase is carried out with a classification task. We elaborately described each module in this section.

4.1 Signal preprocessing

The extracted and recorded EEG signals from subjects during sleep hours are contained in different types of artefacts and noise. In general, the signals are always in the form of a mixture of some noise. On further processing of the input signal for analysis of the irregularities, first of all, we need to reduce the irrelevant noise signal which is mixed with raw signals. In this proposed study, we have obtained two techniques to eliminate noise and artefacts. The general form of the recorded signals is

\[ R_S = E_S + N_S \]  

where \( E_S \) represents the original EEG signal and \( N_S \) represents the noise signal. Our main objective is to reduce the noise from \( E_S \) as close as we could so that we finally obtain an undistorted brain signal information. For this, we have obtained the principal component analysis, which supports reducing the noise problems with the concept of decorrelating the correlated features by identifying the linear relationship between different observations. Similarly, sometimes during recording, the brain signals from subjects, some muscle artefacts, and eye movements occurred and these occurrences are included in the actual brain recordings and it may create sometimes a problem to identify the actual diagnosis. So to eliminate these factors, we have obtained a bandpass Butterworth filter method with a lower cut-off frequency of 0.5 Hz and a higher cut-off frequency of 45 Hz. However, the EEG signal consists of the different characteristics of the wave patterns with different frequency ranges. Generally EEG signals are composed of alpha (\( \alpha \)), theta (\( \theta \)), delta (\( \delta \)), beta (\( \beta \)), sawtooth, sigma (\( \sigma \)), and K-complex characteristic waves. These wave forms appeared during the different stages of sleep. Hence it is important to analyse the different wave forms during the sleep stage classification for identifying the irregularities in the sleep patterns.

The characteristics of the waveform with regard to the different stages of sleep for different frequency ranges are shown in Table 3.

During the sleep time, the \( \sigma \) and \( \delta \) waves occurred in the N2 and REM sleep stages respectively. Similarly, the \( \alpha \) and \( \beta \) waves are more dominant during the Wake and REM sleep stages, and \( \theta \) wave forms are seen in the N1 sleep stage. For that reason, it is important to analyse the characteristics of the EEG signals. Therefore, during the pre-processing the EEG signals are divided into different frequency bands: (0.5–2 Hz), sawtooth (2–6 Hz), (4–8 Hz), (8–13 Hz), (12–14 Hz), and (12–30 Hz) for discriminating the characteristic waveforms of the EEG signals, respectively.

Figure 2a–d presents the sleep behaviour for the subjects 15, 16, 19, and 23 respectively, who were affected with different sleep problems. Similarly, Figure 2e–h presents the behaviour of the healthy control subjects 01, 02, 05, and 08 respectively.
### TABLE 3  EEG signal characteristics for each sleep stage

| Sleep stage | Characteristics of the wave forms | Frequency ranges (Hz.) |
|-------------|-----------------------------------|------------------------|
| W           | $\alpha$ and $\beta$             | $8 \leq \alpha < 13$  |
|             |                                   | $13 \leq \beta < 30$  |
| N1          | $\alpha$ and $\theta$            | $5 \leq \theta < 7$   |
| N2          | $\sigma$ and $k$ – complex        | $12 \leq \sigma < 15$ |
|             |                                   | $0.5 \leq k$ – complex $< 1$ |
| N3          | $\delta$                          | $0.5 \leq \delta \leq 4$ |
| REM         | $\alpha$, $\beta$, $\theta$, and Sawtooth | $2 \leq \text{sawtooth} < 6$ |

Abbreviations: EEG, electroencephalogram; REM, rapid eye movement; W, awake.

**FIGURE 2** Sleep characteristics of each participated subject in this proposed study (a–h). REM, rapid eye movement.
4.2 | Feature extraction

It is one of the important steps for the sleep stage classification because if the suitable features are not selected well then it may put an impact on the performances of the classification task. So it is even more important to design an effective system to extract the features from EEG signals because the EEG signals are like the non-stationary time series data and consist of many numbers of segments that are themselves stationary. In this proposed study, the input EEG signals are segmented into 30 s (6000 sample points).

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

For analysis as a whole, we considered 12 features as key features to represent the sleep behaviour EEG records in this study. It is one of the structured methods and comprehends classification techniques incomparable to another classification algorithm. DT in majority is used by different types of classification tasks [67] and the major cause behind this is its simplicity and ease of understanding the rules regarding the tree structures. A decision tree is constructed from a considered training dataset and each sample of the dataset is contained in the feature values and its class labels. Generally, DT is working like an inductive inference. The major advantage with DT is, it can deal with noisy data and missing data in the dataset. It is also used for multiple stages and consecutive approach during the classification procedure. During the first step of classification, the tree is generated, after which the data is applied one by one to the classification process. Each node in the decision tree is represented as testing features belong to the training set and the generated branch from this node is the proper value of the feature. There are so many algorithms which are designed in connection to DT, but some algorithms are more accepted a subject to different types of classification applications. Some of the most used algorithms in the literature are C4.5, ID3, and C5.

4.3 | Feature selection

It is one of the basic important sections in the machine learning problem; basically its supports for better analysis of any pattern classification-related task. The major advantage of the feature selection is to reduce the complexity and increase the performance of the model. It has been observed from different studies that the selection of appropriate features is a major obstacle generally in the learning problems. Sometimes, it is observed that the original features are not always well-performing for classification tests. During analysis, some features may not be effective for the reason of being irrelevant, redundant and noisy, and sometimes all this nature of features reduces the performance level. Therefore feature selection is a necessity for the task of classification and it indirectly reduces the computation cost [65]. In this study, we used the ReliefF weight feature selection algorithm for screening the relevant features which directly help to recognize the irregularities of the sleep pattern. ReliefF is one of the supervised features of the weighting selection algorithm. It helps to analyse the extent to which features are most useful for discriminating between different stages and measures the effectiveness. The major advantage of this algorithm is to deal with unknown and redundant data [66]. This algorithm is computed using this mathematical expression.

\[
W_T(i) = W_T(i) - \frac{\text{diff}(i, x_k, x_b)}{m} + \frac{\text{diff}(i, x_k, x_b)}{m}
\]

Equation (2) formulated to find out the quality weight \( W_T(i) \) for each feature vector \( x_k \), \( k = 1, 2, 3, ..., F \) and it calculates the nearest miss \( x_m \) and nearest hit \( x_b \) for randomly selected features \( m \).

4.4 | Classification

4.4.1 | Decision tree classifier

It is one of the structured methods and comprehends classification techniques incomparable to another classification algorithm. DT in majority is used by different types of classification tasks [67] and the major cause behind this is its simplicity and ease of understanding the rules regarding the tree structures. A decision tree is constructed from a considered training dataset and each sample of the dataset is contained in the feature values and its class labels. Generally, DT is working like an inductive inference. The major advantage with DT is, it can deal with noisy data and missing data in the dataset. It is also used for multiple stages and consecutive approach during the classification procedure. During the first step of classification, the tree is generated, after which the data is applied one by one to the classification process. Each node in the decision tree is represented as testing features belong to the training set and the generated branch from this node is the proper value of the feature. There are so many algorithms which are designed in connection to DT, but some algorithms are more accepted a subject to different types of classification applications. Some of the most used algorithms in the literature are C4.5, ID3, and C5.
**TABLE 4** Short explanation on extracted features

| Feature no. | Extracted feature equation | Feature descriptions |
|-------------|---------------------------|----------------------|
| 1 | Mean(\(x\)) = \(\frac{1}{N} \sum_{i=1}^{N} x_i\) [69] With \(N\) = the length of the data sample \(x\) and \(\bar{x}\) is the mean of the data sample | The mean electrical potential of an epoch is calculated. It also measures the central tendency in the data points. |
| 2 | Maximum = Max[\(x_i\)] [58] | It is used to quantify the range of data and it helps to find the magnitude of the signal baseline. |
| 3 | Minimum = min[\(x_i\)] [58] | It helps to determine how the data is dispersed with respect to the value of the mean. |
| 4 | Variance (Var) = \(\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2\) [61] With \(N\) = the length of the data sample \(x\) and \(\bar{x}\) is the mean of the data sample | It helps to measure how diversely the dataset spreads out. It computes the difference between the maximum and minimum values, and from this we can get an estimate of the spread of the data. |
| 5 | Range = \((X_{Max} - X_{Min})\) [62] | It helps to determine the most frequent value used to measure the locations of the time series. |
| 6 | Mode = \(L + \frac{(\bmod-(\bmod-1)h)}{(2(f_m - f_{(m-1)}) + (f_m - f_{(m-2)})}}\) [62] \(L\) = lower limit of the modal class \(b\) = size of the class interval \(fn\) = frequency of the modal class \(f1\) = frequency of the class preceding the modal class \(f2\) = frequency of the class succeeding the modal class | It is used to calculate the quantity of variation and dispersion of the data. |
| 7 | Median = \(\frac{X_{\frac{N}{2}}}{\text{rank}}\) [62] With \(N\) = the length of the data sample \(x\) | It helps to get the information about the centre and spread of the signal data. |
| 8 | Q3 = \((N + 1)\times0.75\) [63] With \(N\) = the length of the data sample \(x\) | The quartile analysis provides amplitude information of the signal, which helps to discriminate the sleep stages. It defines the value below which 75% of the random variables data’ is located. |
| 9 | Signal Skewness (skew) = \(\sum_{i=1}^{N} \frac{(x_i - E(x))}{\sigma}\) [41] Where \(N\) is the length of the signal data \(x_i\), \(\sigma\) is the standard deviation of the sample data for all \(i\) \(E(x)\) is the expected mean value \(E(x) = \sum_{i=1}^{N} p(x_i)\cdot p_i\) presents the probabilities with associated to the signal data \(x_i\) | The skewness helps to measure the symmetry of the signal distribution with respect to the mean value. The normal distribution of the signal is zero, while the positive and negative skewness indicate that the data are skewed into the right and left hand sides. It is a higher order statistics measure (third moment) |
| 10 | Signal kurtosis (kurt) = \(\sum_{i=1}^{N} \frac{(x_i - E(x))^4}{\sigma^4}\) [64] Where \(N\) is the length of the signal data \(x_i\), \(\sigma\) is The standard deviation of the sample data for all \(i\), and \(E\) is the expected mean value | It measures whether the data is peaked or flat relative to the normal distribution. It is a higher order statistics measure (fourth moment) |
| 11 | Zero crossing Rate (ZC) = \(\sum_{i=1}^{N} ZC(i)\) [62] | It provides information with regards to the number of instances where the EEG signal crosses the references line. |

**Abbreviation:** EEG, electroencephalogram.

### 4.4.2 K-nearest neighbor classifier

It is one of the mature and simplest theoretically models incomparable to other machine learning classifiers [68]. The main important working style of KNN is to find a similar characteristic between the samples by measuring the distance. It is most acceptable in the case of multi-modal distribution data. It is difficult to decide the boundaries for the different classifiers when the same samples of a certain class are scattered modularly in different locations of the feature vector. But it can manage with the KNN algorithm; it assigns a label to each an input data, and this process is managed by computing the majority of vote of its k-nearest sample points [69]. The major measurement indicator of KNN is to compute the distance between objects located in the feature space, to measure the distance; in general, two mathematical formulae are used named as the Euclidean distance and Manhattan distance.

### 4.4.3 Random forest

This algorithm is proposed by Breimant and is one of the popular classification techniques that use multiple tree structures for training the data and predicting the samples [70].
Each tree requires randomly sampled data values and separate classifiers. The major difference between RF and other classification techniques is that the input is selected in a random manner using bootstrap selection methods. This whole method continues till the noisy and outlier samples do not desensitize and at last, the output is computed by the voting approaches.

4.5 | Model performance evaluation

To validate the proposed model performance which is subjected to the sleep stage classification system, we obtained five evaluation metrics, including, sensitivity [71], specificity [35], precision [72], classification accuracy [73] and F1 Score [74]. The mathematical expression of these statistical indices is defined in Equations (3)–(7).

Apart from above all the evaluation metrics, additionally, we also include the k-cross-validation techniques in our experiment. It is one of the popular measurements in the problem type of the pattern recognition. It is used most of the time in the classification method by dividing the correctly identified classification results by the total number of cases. In this procedure the whole dataset is divided into k subsets, out of which one subset is considered for the testing portion, while others are considered as training subsets. In this work, we have used the 10-fold cross-validation techniques upon the considered dataset.

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \tag{3}
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \tag{4}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \times 100\% \tag{5}
\]

\[
\text{Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \times 100\% \tag{6}
\]

\[
F1\text{Score} = \frac{2 \times TP}{(2TP + FP + FN)} \tag{7}
\]

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively.

5 | RESULT ANALYSIS AND DISCUSSION

To measure the performance of the proposed system, we designed a series of experiments using datasets described in
Section 3. Here it is already mentioned that we have validated the model with the help of different number of statistical features. The whole experiment obtained three different approaches to verify the performance of the model. The first experiment was conducted considering the total extracted features. Similarly, in the second and third approaches, we use the best 9 and 6 features through a feature selection algorithm and considered input for the classification model. All these three approaches are applied to each subject enrolled for this research work. The whole experiment was implemented through the Matlab (Version 2017b) on a computer with memory of 8 GB RAM, 3.40 GHz Intel(R) Core (TM) i7 CPU processor machine.

The proposed sleep study was conducted with one dataset and two different category subject; one category of subjects were affected with sleep-related disorders and the other one completely healthy with no prior controlled medication. We considered the single channel C3-A2 of the EEG signal as an input channel for the proposed model. The sample size of each subject was the same, that is, 750 epochs and the length of each epoch is the 30 s. The sampling frequency is 200 Hz. So for individual subjects, the input size is 750 x 6000.

To analyse the sleep characteristics of each subject and its irregularities which occurred during sleep, we extracted time-series information from the recorded signals. In this work, two different categories of subjects were having symptoms of sleep-related disorders and other categories of subjects belong to completely healthy controls, where we did not find any type of sleep-related problems. In this work, the major objective is to determine which set of combination features gives the best results for sleep stage classification accuracy.

5.1 Set of twelve features

First the experiment is carried out with consideration of 12 features \{X_{ME},X_{VAR}, X_{MAX},X_{MIN}, X_{RAN},X_{ME},X_{MODE},X_{SD}, X_{ZCR},X_{SK},X_{KU},X_{TQ}\}, which were used for classify the sleep stages. From each feature, 750 segments were extracted. Table 5 (A) presented the confusion matrix for sleep-disordered subjects and Table 5 (B) presented for healthy control subjects.

The first experimental results achieved with the help of three classification techniques obtained with the consideration of 12 features are presented in Figure 4a–c for sleep-disordered subjects, similarly, in Figure 5a–c for healthy control subjects. The most-reported results with the input of the category-I subjects who performed best with random forest classification techniques and an overall accuracy resulted from 91.60%, 84.13%, 86.26%, and 88.93% for subject-15, 16, 19, and 23 respectively. Similarly, the performance achieved for healthy control subjects were best with the ensemble classifier, and it has been observed that accuracies for subject-1, subject-5, and subject-8 exceeded 90%. The sensitivity of subject-15 achieved the best and is incomparable to other subjects; the results reported for DT, KNN, and RF are 90.20%, 96.86%, and 94.71% respectively. In the case of healthy subjects, the sensitivity of subject-01 is performed best for all three classifiers. The reported results are 89.10%, 89.10%, and 94.46% for DT, KNN, and RF respectively.

| Table 5 | Confusion matrix presentation for two sleep stages based on 12 features (A) Category-I subjects, (B) Category-II subjects |
|---------|-------------------------------------------------|
| Subject-15 | Wake | Sleep | Subject-16 | Wake | Sleep |
| DT | Wake | 187 | 53 | SVM | Wake | 165 | 86 |
| | Sleep | 50 | 460 | | Sleep | 68 | 431 |
| KNN | Wake | 174 | 66 | KNN | Wake | 156 | 95 |
| | Sleep | 16 | 494 | | Sleep | 29 | 470 |
| RF | Wake | 204 | 36 | RF | Wake | 176 | 75 |
| | Sleep | 27 | 483 | | Sleep | 44 | 455 |

| Subject-19 | Wake | Sleep | Subject-23 | Wake | Sleep |
| DT | Wake | 377 | 63 | DT | Wake | 230 | 52 |
| | Sleep | 61 | 249 | | Sleep | 59 | 409 |
| KNN | Wake | 372 | 63 | KNN | Wake | 214 | 68 |
| | Sleep | 111 | 199 | | Sleep | 36 | 432 |
| RF | Wake | 384 | 56 | RF | Wake | 235 | 47 |
| | Sleep | 47 | 263 | | Sleep | 36 | 432 |

(A)

| Subject Healthy-01 | Wake | Sleep | Subject Healthy-02 | Wake | Sleep |
| DT | Wake | 175 | 52 | SVM | Wake | 189 | 89 |
| | Sleep | 57 | 466 | | Sleep | 76 | 396 |
| KNN | Wake | 175 | 52 | KNN | Wake | 158 | 120 |
| | Sleep | 57 | 466 | | Sleep | 44 | 428 |
| RF | Wake | 191 | 36 | RF | Wake | 209 | 69 |
| | Sleep | 29 | 494 | | Sleep | 48 | 424 |

| Subject Healthy-05 | Wake | Sleep | Subject Healthy-08 | Wake | Sleep |
| SVM | Wake | 375 | 36 | SVM | Wake | 461 | 51 |
| | Sleep | 36 | 303 | | Sleep | 53 | 185 |
| KNN | Wake | 389 | 22 | KNN | Wake | 491 | 21 |
| | Sleep | 25 | 314 | | Sleep | 36 | 202 |
| RF | Wake | 388 | 23 | RF | Wake | 493 | 19 |
| | Sleep | 36 | 303 | | Sleep | 38 | 200 |

(B)
FIGURE 4  Performance statistics results for sleep disordered subjects (a) with decision tree, (b) with K-nearest neighbor, (c) with random forest classifiers

FIGURE 5  Performance statistical results for healthy control subjects (a) with decision tree, (b) with k-nearest neighbor, (c) with random forest classifiers
The second experiment is based on the top nine features, which are selected through the ReliefF feature selection algorithm. The selected top nine features are \{X_{ME}, X_{MIN}, X_{MODE}, X_{ME}, X_{MAX}, X_{RAN}, X_{TQ}, X_{SD}, X_{VAR}\}. From each segment, $10 \times 750$ epochs were extracted. The confusion matrix for

### TABLE 6
Confusion matrix presentation for two sleep stages based on 12 features (A) Category-I subjects, (B) Category-II subjects

|       | Subject-15  | Subject-16  | Subject-17  | Subject-18  | Subject-19  | Subject-20  |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|
| Wake  | 162         | SVM         | Wake        | 226         | Sleep       | Sleep       |
| Sleep | 31          | 9           | Sleep       | 15          | 16          | 34          |
|       | 146         | KNN         | Wake        | 213         | Sleep       | Sleep       |
| Wake  | 170         | RF          | Sleep       | 34          | 16          | 437         |
| Sleep | 116         | RF          | Sleep       | 114         | 55          | 494         |
|       | 309         | SVM         | Sleep       | 65          | 55          | 65          |
| Wake  | 125         | KNN         | Sleep       | 85          | 48          | 48          |
| Sleep | 130         | RF          | Sleep       | 114         | 46          | 46          |

(A)

|       | SubjectH-1  | SubjectH-2  | SubjectH-3  | SubjectH-4  | SubjectH-5  | SubjectH-6  |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|
| Wake  | 165         | SVM         | Wake        | 226         | Sleep       | Sleep       |
| Sleep | 51          | 9           | Sleep       | 15          | 16          | 437         |
|       | 149         | KNN         | Wake        | 213         | Sleep       | Sleep       |
| Wake  | 170         | RF          | Sleep       | 34          | 16          | 492         |
| Sleep | 116         | RF          | Sleep       | 114         | 55          | 494         |
|       | 309         | SVM         | Sleep       | 65          | 55          | 65          |
| Wake  | 125         | KNN         | Sleep       | 85          | 48          | 48          |
| Sleep | 130         | RF          | Sleep       | 114         | 46          | 46          |

(B)

### FIGURE 6
Statistical performance results from sleep disordered subjects based on nine features (a) decision tree classifier (b) k-nearest neighbor classifier (c) random forest classifier

5.2 | Set of Nine Features

The second experiment is based on the top nine features, which are selected through the ReliefF feature selection algorithm. The selected top nine features are $\{X_{ME}, X_{MIN}, X_{MODE}, X_{ME}, X_{MAX}, X_{RAN}, X_{TQ}, X_{SD}, X_{VAR}\}$. From each segment, $10 \times 750$ epochs were extracted. The confusion matrix for
both category subjects based on the nine features is presented in Table 6 (A) and (B).

It can be reported from the experiment that, the achieved performance from both categories of subjects degraded when the features’ number reduced incomparably to the first experiment. The performance results obtained from category-I subjects are presented in Figure 6 and similarly, the results for normal healthy control subjects are presented in Figure 7. The accuracies are best achieved with a combination of nine features for the subject-15 from all the three classification techniques as 90.13%, 90.93%, and 93.20%. Similarly, the best

**TABLE 7** Confusion matrix representation (A) Category-I subjects, (B) Category-II subjects

| Subject-15 | Wake | Sleep | Subject-16 | Wake | Sleep |
|------------|------|-------|------------|------|-------|
| DT         | Wake | 211   | Sleep      | 33   | 474   |
|            | Sleep| 32    |            | 42   | 412   |
| KNN        | Wake | 201   | Sleep      | 12   | 495   |
|            | Sleep| 42    |            | 17   | 437   |
| RF         | Wake | 217   | Sleep      | 20   | 487   |
|            | Sleep| 26    |            | 22   | 432   |
| Subject-19 | Wake | 335   | Sleep      | 38   | 341   |
|            | Sleep| 36    |            | 21   | 441   |
| KNN        | Wake | 555   | Sleep      | 26   | 353   |
|            | Sleep| 16    |            | 18   | 449   |
| RF         | Wake | 348   | Sleep      | 29   | 350   |
|            | Sleep| 23    |            | 14   | 453   |

(A)

| SubjectH-1 | Wake | Sleep | SubjectH-2 | Wake | Sleep |
|-------------|------|-------|-------------|------|-------|
| DT          | Wake | 312   | SVM         | Wake | 262   |
|            | Sleep| 35    | Sleep       | 29   | 431   |
| KNN         | Wake | 324   | KNN         | Wake | 256   |
|            | Sleep| 28    | Sleep       | 26   | 434   |
| RF          | Wake | 332   | RF          | Wake | 260   |
|            | Sleep| 20    | Sleep       | 23   | 437   |
| SubjectH-5 | Wake | Sleep | SubjectH-8 | Wake | Sleep |
| DT          | Wake | 395   | Sleep       | 35   | 295   |
|            | Sleep| 25    | Sleep       | 35   | 196   |
| KNN         | Wake | 393   | Sleep       | 21   | 309   |
|            | Sleep| 27    | Sleep       | 32   | 199   |
| RF          | Wake | 404   | Sleep       | 19   | 311   |
|            | Sleep| 16    | Sleep       | 27   | 204   |

(B)
above 90% for all classifiers and similarly for Category-II subjects, the sensitivity of the subject-I was reported to be above 90% with the RF classifier.

5.3 | Set of five features

The third experiment conducted with a set of five features to both category subjects. The five features were selected based on the weightage of features which was generated from the Relieff feature selection algorithm. The selected five features are \(X_{\text{ME}}, X_{\text{MNN}}, X_{\text{MODE}}, X_{\text{MEN}}, \text{ and } X_{\text{MAX}}\). The vector representation of each segment was represented as \(750 \times 6\). Confusion matrix for category-I subjects are shown in Table 7 (A) and for category-II subjects are shown in Table 7 (B).

The achieved statistical results are shown in Figure 8a–c for subjects who were affected with sleep-related disorders and Figure 9a–c for healthy subjects.

It has been seen that a combination of the best five features is the most effective for the classification of sleep stages for both categories of subjects. The results reported for the subjects who were affected with the sleep disorder reached above 90% for each obtained classifier. The Subject-23 achieved the highest classification accuracy of 93.42%, 93.60%, and 94.03% with decision tree, k-nearest neighbor, and random forest classifiers, respectively.

![Figure 8](image1.png)

**Figure 8** Statistical performance results from sleep disordered subjects based on the five features (a) decision tree classifier (b) k-nearest neighbor classifier (c) random forest classifier.

![Figure 9](image2.png)

**Figure 9** Statistical performance results from normal healthy subjects based on five features (a) decision tree classifier (b) k-nearest neighbor classifier (c) random forest classifier.
FIGURE 10  Accuracy comparison results in between the proposed model with an ensemble classifier in category-I subjects (a) subject-15, (b) subject-16, (c) subject-19 (d) subject-23

FIGURE 11  Comparative results on the classification accuracy in between the proposed model and the ensemble for category-II subjects (a) subject-1, (b) subject-2, (c) subject-5 (d) subject-8
### Table 8
Comparisons of Classification Accuracies (CAs) (in %) of our proposed model with other state-of-the-art techniques which used same features and datasets

| Author | Features | Classifier | Classification accuracy |
|--------|----------|------------|-------------------------|
| Khalighi et al. [75] 2011 | 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 | 95% |
| Hugo Simoes et al. [76] 2010 | Relative spectral power, slow-wave index, Hjorth parameters, entropy, skewness, kurtosis, harmonic parameters | Bayesian classifier | 83% |
| Khalighi et al. [77] 2016 | Entropy and energy, percent of energy, mean, standard deviation, relative spectral power, harmonic parameters, Hjorth parameters, percentile (20,50,75) | SVM | 93.97% |
| Sousa et al. [78] 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. [79] 2013 | Entropy and energy, percent of energy, mean, standard deviation, relative spectral power, harmonic parameters, Hjorth parameters, percentile (20,50,75) | SVM | 81.74% |
| KD Tzimourta et al. [80] 2018 | Energy | RF | 75.29% |
| Najdi et al. [81] 2017 | Multi-domain features | Stacked sparse auto-encoders (SSAE) | 82.3% |
| Kalbkhani, H. et al. [82] 2018 | Entropy features | SVM | 83.33% |
| Proposed | Mean, minimum, median, mode, range, standard deviation, variance, zero crossing rate, signal skewness, maximum, standard deviation, variance, third quartile, signal kurtosis, | DT (Category-I) | 93.42% |
| | | KNN (Category-I) | 94.4% |
| | | RF (Category-I) | 96.70% |

Abbreviations: DT, decision forest; KNN, K-nearest neighbor; RF, random forest.

### Table 9
Comparisons of sleep staging accuracy (%) between the proposed methodology and the recent methods

| Author | Features | Classifier | Classification accuracy |
|--------|----------|------------|-------------------------|
| Dhok, S. et al. [83] 2020 | Wigner-Ville Distribution(WVD) + Renyi entropy features | SVM | 87.45% |
| Huang, W. et al. [84] 2019 | Time and frequency domain features | SVM | 92.04% |
| Sharma, M. et al. [85] 2019 | Log-energy (LE), signal-fractal-dimensions (SFD), and signal-sample-entropy (SSE) from all seven SBs | SVM | 91.5% |
| Memar, P. et al. [86] 2018 | Multiple features | RF | 95.31% |
| Mendez et al. [87] 2016 | Mean, standard deviation, skewness, kurtosis and energy | KNN | 89% |
| Hassan, A.R et al. [39] 2016 | Statistical moment-based features | RF | 90.69% |
| Da Silveira et al. [88] 2016 | Discrete wavelet transform | RF | 91.5% |
| Hassan, A. R. et al. [89] 2015 | Spectral features | DT | 82.03% |
| Proposed | Mean, minimum, median, mode, range, standard Deviation, variance, zero crossing rate, signal skewness, maximum, standard deviation, variance, third quartile, signal kurtosis, | DT (Category-I) dataset | 93.42% |
| | | KNN (Category-I) dataset | 94.4% |
| | | RF (Category-I) dataset | 96.70% |

Abbreviations: DT, decision forest; KNN, K-nearest neighbor; RF, random forest; SVM, support vector machine.
and 96.40%; similarly the healthy subject-5 achieved best classification results and the reported accuracy results are 92.00%, 93.60%, and 95.33% for DT, KNN and RF classifiers respectively. The highest sensitivity results reported for subject-23 is above 95%.

5.4 | Comparative analysis

We used two performance comparison analyses for determining the analysis outcome of the proposed model. The first analysis was compared between the proposed classification results with the ensemble (Bagging) classification performance. The second analysis was completely based on the already existing research works with similar channels or similar datasets (Figure 10).

5.4.1 | Comparison with the ensemble (bagging) classifier

In this section, we have made a comparative analysis in terms of the classification accuracy. The comparison between the ensemble classifier and the results is achieved from the proposed model with a different combination of statistical features. The same combinations of feature sets were used during comparative analysis. It was noticed that the accuracy achieved through the proposed model was better than the ensemble classifier for five combinations of feature sets for category-I subjects, who were affected by the different types of sleep-related disorders. Similarly, for category-2 subjects, who were completely healthy, the proposed model achieved a better performance incomparable to the ensemble classification performance with 12 feature combinations. Figures 10a–d and 11a–d represent the comparative accuracy results with all the three feature combinations for both category-I and category-II subjects.

5.4.2 | Performances comparisons with the existing contributed classification sleep stage methods

The results achieved in the current research work are compared in between different contributed state-of-art literature, in which the obtained input channel is EEG, two stage classification, statistical features, and datasets are addressed. In Table 8, the features used in the proposed research work are compared with the others used by related works using single-channel EEG signal of the ISRUC-Sleep dataset. Comparisons with other similar research proposals are available in the literature which must taken into consideration during the use of the single-channel EEG, and different features and classification models presented in Table 9.

Khalighi et al. [70] used the maximum overlap discrete wavelet transform and obtained both linear and non-linear properties and mRMR feature selection algorithm for screening the suitable features. The system reported an overall accuracy of 95% for the classification of the wake-sleep stages using the SVM classification techniques.

Hugo Simoes et al. [71] used the R-square Pearson correlation coefficient and the selected relevant features were applied into the Bayesian classifier and achieved the overall classification accuracy as 83%.

Khalighi, S. et al. [35] used three categories of subjects records from ISRUC-Sleep repository and extracted both temporal and spectral features extracted from the obtained input channel and applied the SSM4S classification method, thus achieving overall classification accuracy with consideration of the ISRUC-Sleep Subgroup-I, Subgroup-II, and Subgroup-III as 94.10%, 92.40%, and 95.39%, respectively.

Sousa et al. [72] proposed a two-step classifier based on the EEG signal and obtained the SVM classifier for distinguishing epochs suspected of misclassification and as well as both times, and frequency domain features and classified features were forwarded into the SVM classifier with classification accuracy being reported as 86.75%.

Khalighi et al. [73] designed subject-independent improved automated sleep stage classification with application wake-sleep classification and classified through the SVM classifier achieving 81.74% overall classification accuracy.

K D Tzimourta et al. [74] proposed an automated sleep staging system using brain EEG signals. The obtained energy features fed into the random forest classification model. The model reported an overall accuracy of 75.29% for five sleep states classification.

Najdi et al. [75] proposed a sleep study based on the two-layer stacked sparse auto-encoder and obtained frequency, time-frequency, time-domain features which were extracted from EEG signals. The resulted classification accuracy was reported as 82.2%.

Finally, Hashem Kalbkhani [76] introduced the stockwell transform for signal decomposition and the decomposed features were processed through the SVM and KNN classifier. The average accuracy reported for SVM as 82.33% and for KNN as 81.00.

6 | CONCLUSION AND FUTURE WORK

The current research work presents a robust system for sleep stage classification according to sleep rule-based AASM which is a single channel of EEG signals. The proposed methodology was carried out more than the EEG epoch time of 30 s. The data used in the experiment are available on the ISRUC-Sleep database. The basic contribution of this work is the use of statistical features for analysing the sleep characteristics and its three different combination of features used to classify the two-state sleep stages. Three individual experiments with three different sets of features are applied to both sleep-disordered subjects and healthy control subjects. Three machine learning classifiers such as DT, KNN, and RF are obtained for classifying the sleep stages. Additionally, we used the relief feature selection algorithm for screening the best five and nine
features, which help to systematically screen the sleep characteristics. It is observed that the five-set features yield the best performance for both categories of subjects for identifying the sleep irregularities. The reported accuracy for all individual subjects exceeded above 90%. Next to that, 12 feature combinations have given reasonable accuracy results for the two-state sleep stage classification. The performance of the proposed study was analysed in two different manners, first by comparing the results of the proposed model with ensemble classification results, and second, by comparing with eight existing similar contributions. From comparison results, it is found that our proposed model achieved the best classification accuracy. This work can guide clinicians to diagnose accurately and take appropriate decisions for the treatment of different types of sleep-related disorders.

REFERENCES

1. Young, C.P., et al.: A portable wireless online closed-loop seizure controller in freely moving rats. IEEE Trans. Instrum. Meas. 60(2), 513–521 (2011). https://doi.org/10.1109/tim.2010.2050358
2. Liang, S.F., et al.: Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models. IEEE Trans. Instrum. Meas. 61(1), 1649–1657 (2012). https://doi.org/10.1109/tim.2012.2187242
3. Skeldon, A.C., Dijk, D.J., Derks, G.: Mathematical models for sleep-wake dynamics: comparison of the two-process model and a mutual inhibition neuronal model. Plos One. 9(8), e103877 (2014). https://doi.org/10.1371/journal.pone.0103877
4. Pan, S.T., et al.: A transition-constrained discrete hidden Markov model for automatic sleep staging. Biomed. Eng. Online. 11(1), 52 (2012). https://doi.org/10.1186/1475-925x-11-52
5. Rechtschaffen, A., Kales, A.: A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. Public Health Service, US Government Printing Office, Washington DC (1968)
6. Damien Stevens: The international classification of sleep disorders revised. In: Sleep Medicine Secrets, pp. 213–215. Elsevier (1997)
7. Peker, M., Sen, B., Delen, D.: A novel method for automated diagnosis of epilepsy using complex-valued classifiers. IEEE J. Biomed. Health Inform. 20(1), 108–118 (2016). https://doi.org/10.1109/jbhi.2014.2387795
8. Herwig, U., Sarapu, P., Schönfeldt-Lecuona, C.: Using the international 10-20 EEG system positioning of transcranial magnetic stimulation. Brain Topogr. 16(2), 95–99 (2003). https://doi.org/10.1023/b:brat.0000063333.95979.d4
9. Kelly, J.M., Streek, R.E., Bianchi, M.T.: Recent developments in home sleep-monitoring devices. ISRN Neurology, 2012, 1–10 (2012). https://doi.org/10.5402/2012/678794
10. Kellaway, P.: Sleep and epilepsy. Epilepsia. 26(4), S15–S30 (1985). https://doi.org/10.1111/j.1528-1157.1985.tb05720.x
11. Frawan, L., et al.: Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier. Comput. Methods Progr. Biomed. 108(1), 10–19 (2012). https://doi.org/10.1016/j.cmpb.2011.06.005
12. Achermann, P., Taroiki, L.: Human sleep and its regulation. Kosmos. 63(2),173–180 (2011)
13. Sorensen, A.T., Kokaia, M.: Novel approaches to epilepsy treatment. Epilepsia. 54(1), 1–10 (2013)
14. Atoityn, T.N., et al.: Epileptic seizure prediction using CSP and LDA for scalp EEG signals. Comput. Intell. Neurosci. 2017, 1240323 (2017). https://doi.org/10.1155/2017/1240323
15. Norman, R.G., et al.: Interobserver agreement among sleep scorers from different centres in a large dataset. Sleep. 23(7), 901–908 (2000). https://doi.org/10.1093/sleep/23.7.1e
16. Srivivasan, V., Eswaran, C., Sridha, A.N.: Artificial neural network based epileptic detection using time-domain and frequency-domain features. J. Med. Syst. 29(6), 647–660 (2005). https://doi.org/10.1007/s10916-005-6133-1
17. Temko, A., et al.: EEG-based neonatal seizure detection with Support Vector Machines. Clin. Neurophysiol. 122(3), 464–473 (2011). https://doi.org/10.1016/j.clinph.2010.06.034
18. Banch, C., Pompe, B.: Permutation entropy: a natural complexity measure for time series. Phys. Rev. Lett. 88(17) (2002). https://doi.org/10.1103/physrevlett.88.174102
19. Kannathal, N., et al.: Entropies for detection of epilepsy in EEG. Comput. Methods Progr. Biomed. 80(3), 187–194 (2005). https://doi.org/10.1016/j.cmpb.2005.06.012
20. Bao, F.S., Li, DYC., Zhang, Y.: A new approach to automated epileptic diagnosis using EEG and probabilistic neural network. In: Proceedings of the 2008 29th IEEE International Conference on Tools with Artificial Intelligence. (2008). https://doi.org/10.1109/icai.2008.99
21. Carrozzetti, M., Accardo, A., Bouquet, F.: Analysis of sleep-stage characteristics in full-term newborns by means of spectral and fractal parameters. Sleep. 27(7), 1384–1393 (2004). https://doi.org/10.1093/sleep/27.7.1384
22. Siddiqui, M.M., Srivastava, G., Saced, S.H.: Diagnosis of insomnia sleep disorder using short time frequency analysis of PSD approach applied on EEG signal using channel ROC-LOC. Sleep. Science. 9(3), 186–191 (2016). https://doi.org/10.1016/j.sleep.2016.07.002
23. Kaffashi, F., et al.: An analysis of the kangaroo care intervention using neonatal EEG complexity: a preliminary study. Clin. Neurophys. 124(2), 238–246 (2013). https://doi.org/10.1016/j.clinph.2012.06.021
24. Seber, M.S., et al.: Computer classification of sleep in preterm and full-term neonates at similar postconceptional term ages. Sleep. 19(1), 18–25 (1996). https://doi.org/10.1093/sleep/19.1.18
25. Pirayatska, A., et al.: Automated detection of neonate EEG sleep stages. Comput. Methods Progr. Biomed. 95(3), 31–46 (2009). https://doi.org/10.1016/j.cmpb.2009.01.006
26. Koolen, N., et al.: Automated classification of neonatal sleep states using EEG. Clin. Neurophysiol. 128(6), 1100–1108 (2017). https://doi.org/10.1016/j.clinph.2017.02.025
27. Estévez, P.A.: Polysomnographic pattern recognition for automated classification of sleep-wake states in infants. Med. Biol. Eng. Comput. 40(1), 105–113 (2002). https://doi.org/10.1007/bf02347703
28. Flexer, A., Gruber, G., Dorfner, G.: A reliable probabilistic sleep stage based on a single EEG signal. Artif. Intell. Med. 33(3), 199–207 (2005). https://doi.org/10.1016/j.artmed.2004.04.004
29. Roebuck, A., et al.: A review of signals used in sleep analysis. Physiol. Meas. 35(1), R1 (2014). https://doi.org/10.1088/0967-3334/35/1/r1
30. Chapotot, F., Becq, G.: Automated sleep-wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules. Int. J. Adapt. Contr. Signal Process. 24, 409–423 (2009). https://doi.org/10.1002/acs.1147
31. Sinha, R.K.: Artificial neural network and wavelet based automated detection of sleep spindles, REM sleep and wake states. J. Med. Syst. 32(4), 291–299 (2008). https://doi.org/10.1007/s10916-008-9134-z
32. Ebrahimi, F., et al.: Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients. In: Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2008). https://doi.org/10.1109/iembs.2008.4649365
33. Ronzhina, M., et al.: Sleep scoring using artificial neural networks. Sleep. Med. Rev. 16(3), 251–263 (2012). https://doi.org/10.1016/j.smrv.2011.06.003
34. Liang, S.F., et al.: Development of a human-computer collaborative sleep scoring system for polysomnography recordings. Plos One. 14(7) (2019). https://doi.org/10.1371/journal.pone.0218948
35. Hsu, Y.L., et al.: Automatic sleep stage recurrent neural classifier using energy features of EEG signals. Neurocomputing, 104, 105–114 (2013). https://doi.org/10.1016/j.neucom.2012.11.003
36. Braun, E.T., et al.: Sleep stages classification using spectral based statistical moments as features. Revista de Informática Teórica e Aplicada. 25(1), 11 (2018). https://doi.org/10.22456/2175-2745.74030

37. Zhu, G., Li, Y., Wen, P.P.: 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 (2014). https://doi.org/10.1109/jbhi.2014.2303991

38. Abolayon, K.A.I., Ocbagabir, H.T., Eazipour, M.: Efficient sleep stage classification based on EEG signals. In: Proceedings of the IEEE Long Island Systems, Applications and Technology (LISAT) Conference (2014)

39. Hassan, A.R., Bhuiyan, M.I.H.: Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. Biomed. Signal Process. Contr. 24, 1–10 (2016). https://doi.org/10.1016/j.bspc.2015.09.002

40. Hassan, A.R., Hassan Bhuiyan, M.L.: Automatic identification of sleep states classification based on EEG signals. In: Proceedings of the IEEE Engineering in Medicine and Biology Society, 1876–1880 (2014)

41. Hassan, A.R., Bhuiyan, M.I.H.: A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. J. Neurosci. Methods. 271, 107–118 (2016). https://doi.org/10.1016/j.jneumeth.2016.07.012

42. Hassan, A.R., Bhuiyan, M.I.H.: An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting. Neurocomputing. 219, 76–87 (2017). https://doi.org/10.1016/j.neucom.2016.09.011

43. Hassan, A.R., Bashai, A.: A decision support system for automated identification of sleep stages from single-channel EEG signals. Knowl. Base Syst. 128, 115–124, https://doi.org/10.1016/j.knosys.2017.05.005

44. Kim, J.: A comparative study on classification methods of sleep stages by using EEG. J. Korea Multimed. Soc. 17(2), 113–123 (2014). https://doi.org/10.9717/kcms.2014.17.2.113

45. Shekhikhvand, S., et al.: Sleep Stage Scoring of Single-Channel EEG Signal based on RUSBoost Classifier. In: 2018 IEEE International Conference on Artificial Intelligence in Engineering and Technology (ICAIEET) (2018). https://doi.org/10.1109/icaieet.2018.8638465

46. Akin, M., et al.: Estimating vigilance level by using EEG and EMG signals. Neural Comput. Appl. 17(3), 227–236 (2008). https://doi.org/10.1007/s00521-007-0917-7

47. Ge, Z., Sun, Y.: Sleep stages classification using neural networks with multi-channel neural data. In: Guo Y. et al. (eds) Brain Informatics and Health, BIH 2015. Lecture Notes in Computer Science, 9250 (2015). https://doi.org/10.1007/978-3-319-45033-280

48. Zhang, Y., et al.: A comparison study on multi domain EEG features for sleep stage classification. Comput. Intell. Neurosci., 2017, 1–8 (2017). https://doi.org/10.1155/2017/4574079

49. Diiykh, M., Li, Y.: Complex networks approach for EEG signal sleep stages classification. Expert Syst. Appl. 63, 241–248 (2016). https://doi.org/10.1016/j.eswa.2016.07.004

50. Zhang, B., et al.: EEG-based automatic sleep staging using ontology and weighting feature analysis. Comput. Math. Meth. Med., 2018, 1–16 (2018). https://doi.org/10.1155/2018/6534041

51. Mahvash Mohammadi, S., et al.: Improving time-frequency domain sleep EEG classification via singular spectrum analysis. J. Neurosci. Meth. 273, 96–106 (2016). https://doi.org/10.1016/j.jneumeth.2016.08.008

52. Abolayon, K., et al.: Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation. Entropy. 18(9), 272 (2016). https://doi.org/10.3390/e18090272

53. Burjak, N., et al.: Approximate entropy in the electroencephalogram during wake and sleep. Clin. EEG Neurosci. 36(1), 21–24 (2005). https://doi.org/10.1177/155055940503600106

54. Obaya, M., AbouChadi, F.E.Z.: Automatic classification of sleep stages using EEG records based on fuzzy c-means (FCM) algorithm. In: Proceedings of the 31st National Radio Science Conference (NRSC), 265–272 (2014)

55. Frazan, L., et al.: Classification of sleep stages using multi-wavelet time frequency entropy and LDA. Methods Inf. Med. 49(3), 230–237 (2010). https://doi.org/10.3414/ME09-01-0054

56. Herrera, L.J., et al.: Combination of heterogeneous EEG feature extraction methods and stacked sequential learning for sleep stage classification. Int. J. Neur. Syst. 23(03), 1350012 (2013). https://doi.org/10.1142/S0129065713500123

57. Radha, M., et al.: Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal. In: Proceedings of the IEEE Engineering in Medicine and Biology Society, 1876–1880 (2014)

58. Jo, H.G., et al.: Genetic fuzzy classifier for sleep stage identification. Comput. Biol. Med. 40(7), 629–634 (2010). https://doi.org/10.1016/j.jneumeth.2010.04.007

59. Mendez, M.O., et al.: Sleep staging classification based on HRV-time-variant analysis. In: Proceedings of the International Conference on IEEE Engineering in Medicine and Biology Society, 9–12 (2009)

60. Diiykh, M., Li, Y., Wen, P.: EEG sleep stages classification based on time domain features and structural graph similarity. IEEE Trans. Neural Syst. Rehabil. Eng. 24(11), 1159–1168 (2016). https://doi.org/10.1109/tnsre.2016.2552539

61. Prieto, M.D., et al.: Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. IEEE Trans. Ind. Electron. 60(8), 3398–3407 (2013). https://doi.org/10.1109/tie.2012.2219383

62. Lajeuf, T., et al.: Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. J. Neurosci. Methods. 250, 94–105 (2015). https://doi.org/10.1016/j.jneumeth.2015.01.022

63. Susmáková, K., Krátková, A.: Discrimination ability of individual measures used in sleep stages classification. Artif. Intell. Med. 44(3), 261–277 (2008)

64. Cao, B., et al.: Feature selection in a kernel space. In: Proceedings of the 24th Annual International Conference on Machine Learning, 121–128 (2007)

65. Kononenko, I.: Estimating attributes: analysis and extensions of RE-LIEF. In: Bergadano F., De Raedt L. (eds) Machine Learning: ECML-94. ECML. 1994. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), 784 (1994). https://doi.org/10.1007/3-540-57868-4_57

66. Sen, B., Uçar, E., Delen, D.: Predicting and analyzing secondary education placement-test scores: a data mining approach. Expert Syst. Appl. Int. J. 39, 9468–9476 (2012)

67. Zhang, S., Li, X., Zong, M.: Learning for kNN classification. ACM Trans. Intell. Syst. Technol. 8(31–19), 43 (2017). https://doi.org/10.1145/2995058

68. Fraley, C., Raftery, A.E.: Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. 97(458), 611–631 (2002). https://doi.org/10.1198/016214502760047313

69. Shabani, F., Kumar, L., Ahmadi, M.: A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. Ecol. Evol. 6(16), 5973–5986 (2016). https://doi.org/10.1002/ece3.2332

70. Bajaj, V., Pachori, R.B.: Automatic classification of sleep stages based on the time-frequency image of EEG signals. Comput. Methods Progr. Biomed. 112(3), 320–328 (2013). https://doi.org/10.1016/j.cmpb.2013.07.006

71. Yildiz, A., et al.: Application of adaptive neuro-fuzzy inference system for vigilance level estimation by using wavelet-entropy feature extraction. Expert Syst. Appl. 36(4), 7390–7399 (2009). https://doi.org/10.1016/j.eswa.2008.09.003

72. Sanders, T.H., McCurry, M., Clements, M.A.: Sleep stage classification with cross frequency coupling. In: Proceedings of the 36th Annual International Conference IEEE Engineering in Medicine and Biology Society. (EMBC), pp. 4579–4582 (2014)
74. Powers, D.M.: Evaluation: From Precision, Recall and F-measure to Roc, Informedness, Markedness and Correlation. J. Mach. Learn. Technol. 2, 2229–3981 (2011). https://doi.org/10.9735/2229-3981
75. Khalighi, S., et al.: Efficient feature selection for sleep staging based on maximal overlap discrete wavelet transform and SVM. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2011)
76. Simões, H., et al.: Feature extraction and selection for automatic sleep staging using EEG. In: Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics. 3, 128–133 (2010). https://doi.org/10.5220/0002950601280133
77. Khalighi, S., et al.: ISRUC-Sleep: a comprehensive public dataset for sleep researchers. Comput. Methods Progr. Biomed. 124, 180–192 (2016)
78. Sousa, T., et al.: A two-step automatic sleep stage classification method with dubious range detection. Comput. Biol. Med. 59, 42–53 (2015)
79. Khalighi, S., et al.: Automatic sleep staging: a computer assisted approach for optimal combination of features and polysomnographic channels. Expert Syst. Appl. 40(17), 7046–7059 (2013)
80. Tzimourta, K.D., et al.: EEG-based automatic sleep stage classification. Biomed. J. Sci. &Tech. Res. 7(4) (2018)
81. Najdi, S., Gharbali, A.A., Fonseca, J.M.: Feature transformation based on stacked sparse autoencoders for sleep stage classification. In: Technological Innovation for Smart Systems, 191–200 (2017)
82. Kalbkhani, H., Ghasemzadeh, P.G., Shayesteh, M.: Sleep stages classification from EEG signal based on stockwell transform. IET Signal Process. 13(2), 242–252 (2018). https://doi.org/10.1049/iet-spr.2018.5032
83. Dhok, S., et al.: Automated phase classification in cyclic alternating patterns in sleep stages using Wigner-Ville Distribution based features. Comput. Biol. Med. 119, 103691 (2020)
84. Huang, W., et al.: Sleep staging algorithm based on multichannel data adding and multifeature screening. Comput. Methods Progr. Biomed. 187, 105253 (2019). https://doi.org/10.1016/j.cmpb.2019.105253
85. Sharma, M., et al.: Automated detection of sleep stages using energy-localized orthogonal wavelet filter banks. Arabian J. Sci. Eng. (2019). https://doi.org/10.1007/s13369-019-04197-8
86. Memar, P., Faradji, F.: A novel multi-class EEG-based sleep stage classification system. IEEE Trans. Neural Syst. Rehabil. Eng. 26(1), 84–95 (2018). https://doi.org/10.1109/TNSRE.2017.2776149
87. Mendez, M.O., et al.: Analysis of A-phase transitions during the cyclic alternating pattern under normal sleep. Med. Biol. Eng. Comput. 54(1), 133–148 (2016)
88. Da Silveira, T.L.T., Kozakevicius, A.J., Rodrigues, C.R.: 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 (2016)
89. Hassan, A.R., Bashar, S.K., Bhuiyan, M.I.H.: Automatic classification of sleep stages from single-channel electroencephalogram. In: Proceedings of the Annual IEEE India Conference (INDICON) (2015)

How to cite this article: Satapathy S, Loganathan D, Kondaveeti HK, Rath RK. Performance analysis of machine learning algorithms on automated sleep staging feature sets. CAAI Trans. Intell. Technol. 2021;6:155–174. https://doi.org/10.1049/cit2.12042