Journal of Education for Pure Science
Vol.15, No.2 (2025)
DOI: https://doi.org/10.32792/jeps.v15i2.571
Website: jceps.utq.edu.iq Email: jceps@eps.utq.edu.iq

DOI:

An Intelligent Model for EEG Sleep Stages Classification Using Wavelet Transform Based Hybrid Features Extraction Model

Afrah S. Jabbar[1,*] and Wessam Al-Salman[1]

Department of Computer Sciences, College of Education for Pure Science, University of Thi-Qar, Nasiriyah, 64001, Iraq

[*] Corresponding email: afrahsattar.comp@utq.edu.iq

Received 19/ 10 / 2024, Accepted 14 / 11 /2024, Published 1 / 6 /2025

Abstract:

Background: High-quality sleep plays a major role in improving brain development and lifestyle. Electroencephalogram (EEG) signals are the most important signals collected by polysomnography (PSG) used for sleep staging. Manual sleep scoring is a difficult task, however, developing an automatic sleep stage is important to support experts to detect sleep disorders in early stage. ***Method***: In this paper, an automatic single-channel EEG signal sleep stages classification model is proposed. A Discrete Wavelet Transform (DWT) based EEG feature is suggested. Three types of features including entropy, linear, and statistical features are extracted and evaluated to score sleep stages. First, we applied the DWT to each 30-second epoch to decompose the signal into five bands. Then, EEG features are extracted from each band. EEG signals from two datasets named Dreams, and EFD sleep are used to evaluate the proposed model. ***Results***: We interpreted the results using essential statistical criteria. The results showed that the use of combination features improves the sleep classification results. Based on the results, with the Dream dataset, the classification accuracy rate, Kappa coefficient, and F-score were found 0.97,0.92, 0.95 For the second database, we obtained 0.95,0.94, 0.93 for accuracy rate, Kappa coefficient and F-score respectively. ***Conclusions***: We developed a method to score sleep stages that can be used by healthcare providers to identify sleep disorders.

Keywords: *DWT, EEG, Epoch, classification, sleep stages*.

1-Introduction

Sleep is a complex and active process that involves activities from different neurons. Sleep is an essential phenomenon that maintains learning and memory functions as well as protects metabolizable energy. Sleep experts classified sleep into main two stages, non- rapid eye movement (NREM) and rapid eye movement (REM). The REM stage in which dreams occur, lasts 5-20 minutes at each 90 minutes' interval. While during NREM sleep, neurons become less active, and metabolic rate decreases, as well as blood pressure, and heart rate.

Sleep disorders are among the leading issues that affect the quality of life as sleep disorders cause several life-threatening diseases. Mainly, sleep disorders can be diagnosed using the polysomnography (PSG) tool which is a collection of cardiorespiratory, and neurophysiologic recordings. In clinical units, expert manually scoring PSG. However, this process is tiring and time-consuming. In 1968, the first sleep scoring rule and technology was introduced by Rechtschaffen and Kales (R & K). Then, in 2007, this technique was modified by the American Academy of Sleep Medicine (AASM), who's published a new update as a second version in 2012. Based on the lates update carried out by the AASM, the sleep stages divided into wake (W), stage 1 (N1), stage 2 (N2), stage 3 (N3), and REM, in which the S4 was removed from the rule list of sleep stages. The AASM rules are split whole night sleep into epochs of 30-s- intervals, then, each epoch is assigned to one of sleep stages in the literature, EEG signal is the most important PSG signal that is commonly used for sleep scoring compared with EOG, ECG. A High performance can be achieved by using this signal; however, it is not always easy to extract the most effective features from this signal. Recently, Goldberger A.L et al., [1] proposed hand engineered features-based EEG model. In that study, they adopted frequency, time, and time-frequency domain EEG features. The extracted features were sent into Support Vector Machines (SVM), Random Forests (RF), and Neural Networks (NNs). Shao et al., [2] designed a hybrid intelligent model integrating data intelligence and knowledge intelligence. Cheng et al., [3] Utilised sleep classification model based on a generative adversarial network ensemble deep learning model to address the imbalance issue in sleep scoring. Abdulla et al., [4] designed a multi-channel spectrum-based image model to score EEG sleep. Each EEG segment was converted into an image using a time-frequency image. The diffident texture features were extracted using spectrum pattern analysis.

Other researchers employed transformation techniques and graph-based models. For example, Guo et al. [5] adopted a relative wavelet energy coupled with an artificial neural network. Jo et al. [6] employed a genetic algorithm based on a fuzzy classifier. In that study, the fast Fourier transform with a hanning window was applied to decompose EEG signals. Doroshenkov et al. [7] applied a Markov model based on a fast Fourier transform filter. Various features were extracted and then classified. More recently, graphs-based model has been designed to classify sleep stages. Zhu et al. [8] identified EEG sleep stages using visibility graphs and horizontal visibility graphs. A support vector machine was used to classify graph features into the six sleep stages. Shi et al. [9] suggested multi-a collaborative representation model based on a multi-learning algorithm. A k-means classifier and dictionary learning method were employed in classification phase.

In this study, we used two different databases to evaluate the proposed model including a total of 17,758 epochs extracted from 28 subjects. Single-channel EEG was adopted to score sleep stages. The wavelet-based model was designed to extract features from each 30s epoch of the EEG signals. Linear, statistical, and entropy features were

extracted and evaluated. All EEG recordings were normalized, and then we split the data into training, and testing sets using 10-fold cross-validation. The extracted features were sent into different ensemble and individual classifiers. The results were interpreted using statistical measures including Precision, F-score, and Kappa. Despite the disadvantage of working on single-feature EEG signals, the obtained results prove the success of using the multifeatured model for sleep stages.
In this paper, and we designed a sleep stages classification model based on DWT and entropy features. We have combined different entropy features. Based on our knowledge, this approach has not been used in sleep stages classification before.

2. Experimental EEG recordings

We used two EEG datasets to evaluate the proposed model to identify sleep stages from EEG signals. We gave a brief description of the datasets used in this paper.

2.1. Dataset_1 (Sleep-EDF dataset)

The first EEG dataset was collected from PysioNet [10][11][12][13]. The Sleep-EDF datasets is free available. Each polysmnographic (PSG) recording is included one EOG signal, two EEG signals, one EMG signal, Resporonasal, Eventmarker, EMG Submenta, and Tempbody. EEG signals from 61 subjects were collected. In this paper, we selected 13 subjects randomly and employed those EEG recordings for the evaluation. We selected EEG signals from Pz-Oz channel. The datasets were collected in the period of 1987 to 1994. Different subjects from Caucasian males and females were involved. European data format (EDF) technology was used to store EEG recordings. A frequency of 100 Hz was adopted to sample the original EEG signals. The R& K criteria was employed to segment EEG recordings into epoch of 30 seconds (3000 data points) [14]. All segments were then named as AWA, S1, S2, S3, S4, REM. Table 1 shows the distribution of segments that were used in this study.

2.2. Dataset_2 (St. Vincent's University)

Another dataset from at St. Vincent's University Hospital was used for further evaluation of the proposed model. A period of 6 months was spent to record the dataset from different subjects (Goldberger et al., 2000). In this paper, we selected all subjects for evaluation. The demographic information of subjects was recorded as follows: age 50±10 years, 21 males and 4 females, weight range 25.1-42.5 kg. The PSG was included 2 EOG channels, 2 EEG channels (C3-A2 and C4-A1), and 1 EMG channel. We selected the C3-A2 EEG channel. EEG signals were divided into intervals of 30 seconds (3000 data points).

3. The proposed Method for EEG sleep classification

In this paper, EEG sleep classification model is presented using DWT based on hybrid feature extraction model. Each epoch of 30 second is passed through DWT. As a result, each EEG segment was decomposed into five bands. To reduce the dimensionality of each band, a set of features is extracted. In this paper, three types of features are extracted and investigated including entropy, linear, and statistical features. The extracted features are analysed and selected using statistical metrics. The sleeted features are sent to assemble machine learning models to classify the characteristics of the EEG into five sleep stages. Fig 4 depicts the methodology of the proposed model.

3.1 Discrete wavelets transform (DWT)

DWT is the most effective model to analyse nonstationary signals such as EEG signals. We employed the discrete Wavelet transform (DWT) to analyse EEG signals. The DWT of signal x can be defined [15]as

$$DWT(i,j) = \frac{1}{\sqrt{|2^i|}} \int_{\infty}^{\infty} x(f)\psi\left(\frac{t-2^i k}{2^i}\right) \qquad (1)$$

where $DWT\ (i,j)$ refers to the wavelet coefficients, $x(f)$ denoted to the EEG signal, $\psi(.)$ is a wavelet function, and $2^i\ and\ \ k\ are$ scaling factors.

The EEG signals are decomposed into different frequency bands. A set of approximations and details were generated after passing the EEG signal via a series of high and low-pass filters.

Figure 2 shows an EEG epoch is being anlysed into a set of approximations and details. At each iteration, we employed two digital filters and two down-sampled outputs are employed. We obtained the detail (D1) and the approximation (A1), at the first level. For further decomposition, the same process can be performed for the approximation A1. This process is repeated to obtain the desired output. To identify the appropriate number of the decomposition level as well as the type of wavelet, we conducted several experiments in this paper. We found that the Daubechies (db5) delivered good results compared to other wavelet functions. The five sub-bands were denoted as D1, D2, D3, D4, D5 A5 refers to the decomposition approximation coefficients and D1-D5 are the decomposition detail coefficients. It was observed that the six-level wavelet decomposition and Daubechies (db6) yielded better results compared to others. Therefore, in this study, D5 was chosen empirically.

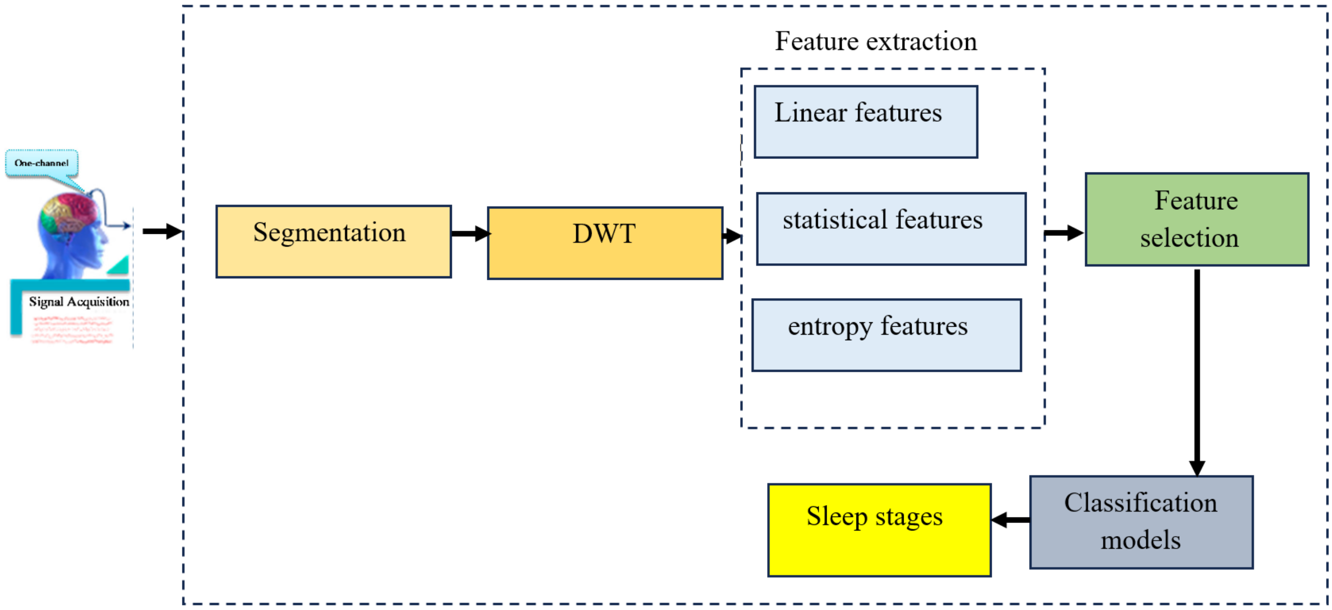

Fig. 1. Methodology of EEG sleep identification

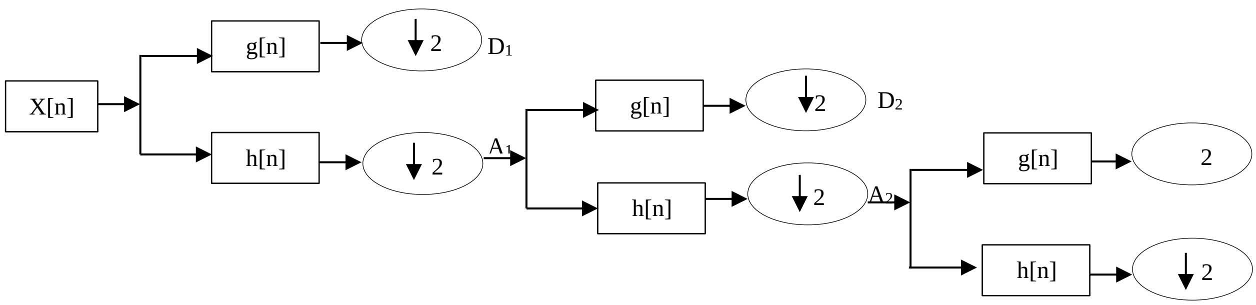

Fig. 2. EEG epoch is being decomposed into three levels.

3.2 Features extraction

Feature extraction is the most important phase of the EEG sleep classification. When the features are not selected well the classification rate can be degraded. As a result, it is an important aspect to design an effective model to pull out the most effective features from EEG signals. In this work, the EEG feature extraction model is designed based on statistical and entropy features. After the decomposition of EEG signals into five bands, each band was passed to the feature extraction model to extract EEG features. Figure 3 describes the feature extraction model. We extract 12 statistical features and 3 entropy features from each of the five bands ($\delta, \ \theta, \ \alpha, \ \beta, \ \gamma$). A Total (5 x 13) feature vectors were extracted from each EEG segment. As a result, each EEG signal is represented as a matrix of NXN (5X13), where N refers to the number of segments for each EEG signal. The features are tested to evaluate their performance. All the features were evaluated using statistical metrics.

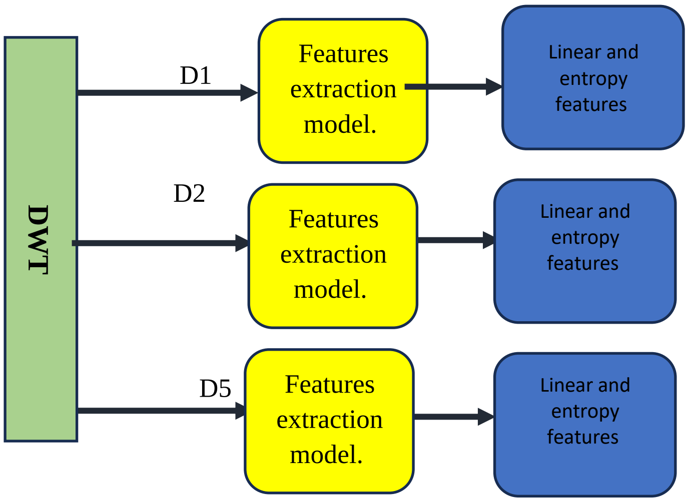

Fig. 3. EEG epoch is being decomposed into three

3.2.1 Statistical features

Based on previous studies, EEG signals exhibit symmetric and skewed behaviour. For symmetric behaviour, we employed the mean and the standard division to achieve that. For the skewed distribution, we employed the following feature {median, range and quartile} to measure the centre of data. We utilised other statistical features, such as minimum, variation, skewness and kurtosis to figure out the important information about EEG signals. A total 12 features of *{median, second quartile, standard deviation, maximum, minimum, mean, mode, range, first quartile, variation, skewness, kurtosis}* were extracted to represent EEG signals.

Table 1 provides a short explanation of the statistical features. The 12 features are denoted as $\{X_1, X_2, X_3, \ldots, X_{12}\}$.

TABLE 1

SHORT EXPLANATION OF THE STATISTICAL FEATURES

No.	Feature name	Formula	No.	Feature name	Formula
1	Maximum	X_{Max}=Max[x_n]	7	Minimum	X_{Min}=min[x_n]
2	Mean	$X_{Mean} = \frac{1}{n}\sum_{1}^{n} x_i$	8	Mode	$X_{Mod} = L + \left(\frac{f_1 - f_0}{2f_1 - f_2}\right) Xh$
3	Median	$X_{Me} = (\frac{N+1}{2})^{th}$	9	Range	$X_{Rang}=X_{Max}-X_{Min}$
4	First Quartile	$X_{Q1} = \frac{1}{4(N+1)}$	10	Standard Deviation	$X_{SD} = \sqrt{\sum_{n=1}^{N}(x_n - AM)\frac{2}{n-1}}$
5	variation	$X_{Var} = \sum_{n=1}^{N}(x_n - AM)\frac{2}{N-1}$	11	Skewness	$X_{Ske} = \sum_{n=1}^{N}(x_n - AM)\frac{3}{(N-1)SD^3}$
6	Kurtosis	$X_{Ku} = \sum_{n=1}^{N}(x_n - AM)\frac{4}{(N-1)SD^4}$	12	Second Quartile	$X_{Q2} = \frac{4}{4(N+1)}$

where X_n=1,2,3....n is a time series, N is the number of data points, AM is the mean of the sample.

3.2.2 Linear features

In addition, we also extract nonlinear features such as Shannon entropy, dispersion entropy, and approximate entropy [9].

1. *Approximate entropy: is a feature that is used to measure the amount of regularity and unpredictability of a signal.*
2. *Shannon entropy*: is applied to measure the degree of uncertainty of a random time series. The larger value of Shannon entropy means more randomness and uncertainty of the timeseries Shannon entropy can be defined as:

$$H = -\sum_{i=1}^{n} S_i \ln(S_i) \quad (2)$$

Where S_i is the probability of the i sample in the timeseries value and H is the Shannon entropy.

3. *Dispersion entropy: is employed to assess the complexity or irregularity of signals* [16].

$$DispEn(m,c) = -\sum_{i=1}^{c^m} p[\,y_m(i)]\log(p[\,y_m(i)]) \quad (3)$$

3.3 Feature selection model based on statistical metrics

In this paper, a features selection model integrated Student t-test, and Wilcoxon test is designed. The student's t-test approach is based on the significance between two samples. The features with a higher t-value are selected. The second metric is Wilcoxon test which is a non-parametric test. It determines whether the average of two samples vary or not based on the normally distributed.

To enhance the sleep stages classification accuracy, the noisy EEG features are removed in this paper using statistical metrics. Our results showed that not all features are important to classify EEG sleep stages. Tables 2, and 3 report the results of feature selection. We can observe that some features did not pass the test. We applied the t-test which works according to the following hypothesis.

$$H_0: \mu_1 = \mu_2 \; vs \; H_1: \mu_1 \neq \mu_2$$

where μ_1 refers to the average of the first sample, μ_2 denotes to average of the second sample, and the level of significance α=0.05. When the p-value ≤0.05 the features are accepted, while the features with values p-value ≥0.05 are rejected. Based on the t-test, we found that the statistical features *{range variance, standard deviation, mean, max, median, and kurtosis}* can be used to differentiate among sleep stages. For entropy features, it was found that the features *{ Shannon entropy and approximation}* are accepted and passed the test.

In this paper, we also employed the Wilcoxon test to select the most appropriate features to classify EEG sleep stages.

Based on Wilcoxon test the following criteria are adopted H_0: refers to features that are belonged to the same population *vs* H_1: refers to the features that not belonged to the same population. We considered that the features with p ≤0.05 are accepted, while the features with p ≥0.05 are considered to be not significant. As a result, not features are accepted to represent EEG sleep stages. The new results based on Wilcoxon test are presented in Table 3. The results of Wilcoxon are compatible and matched with t-tes results.

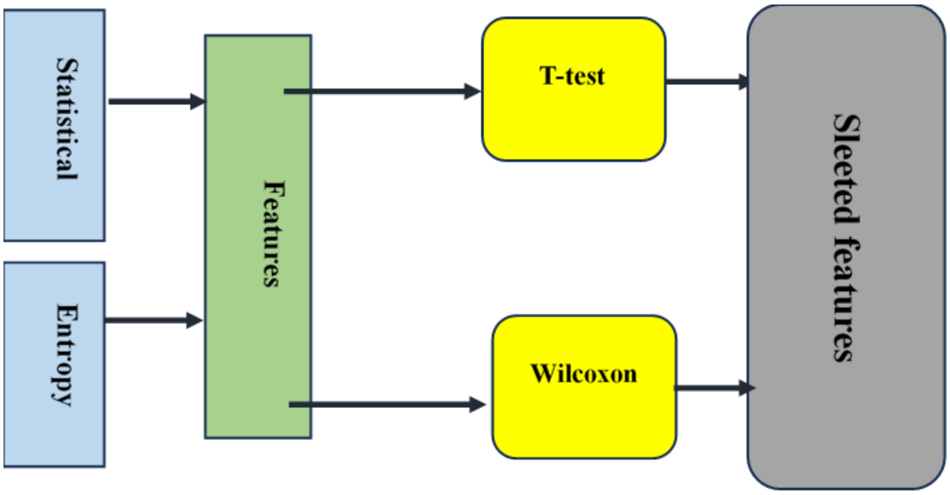

Fig. 4. EEG features selection model

Table 2

Statistical feature evaluation using t-test

Stage	Max	Min	Range	Std	Mean	Median	skewness	Kurtosis	First Q	Second Q	mode	Third Q
AW	0.0111	0.0177	0.0176	0.031	0.0181	0.0179	0.0121	0.0117	0.5926	0.6119	0.6290	0.5822
S1	0.0420	0.0287	0.2020	0.032	0.1120	0.0132	0.0228	0.0109	0.6210	0.6212	0.5311	0.7221
S2	0.0130	0.0034	0.0032	0.056	0.0011	0.0370	0.0216	0.00011	0.6230	0.7243	0.5429	0.6220
S3	0.0230	0.0221	0.0305	0.021	0.0529	0.1101	0.0353	0.0224	0.4922	0.4120	0.546	0.5821
REM	0.0321	0.0372	0.0540	0.041	0.122	0.0117	0.0461	0.0176	0.5621	0.6148	0.5442	0.7221
Statistical feature evaluation using Wilcoxon												
AW	0.0214	0.0245	0.0212	0.012	0.0121	0.0032	0.0104	0.01061	0.5324	0.7612	0.6876	0.5656
S1	0.0321	0.0112	0.0213	0.021	0.1321	0.0103	0.0205	0.01321	0.6245	0.5678	0.5789	0.7789
S2	0.0217	0.0134	0.0311	0.012	0.0453	0.0`12	0.0205	0.00431	0.7543	0.7123	0.7855	0.6767
S3	0.0132	0.0151	0.0214	0.002	0.0532	0.1211	0.0312	0.02065	0.5642	0.5431	0.6556	0.5876
REM	0.0213	0.0121	0.0213	0.011	0.1210	0.0100	0.0404	0.01021	0.5875	0.7863	0.6402	0.7655

Table 3

Entropy feature evaluation using Wilcoxon

Stage	Appr. entropy	Desp.entropy	Shan.entropy
AW	0.101	0.5301	0.0380
S1	0.1221	0.6038	0.1109
S2	0.0341	0.5503	0.0285
S3	0.0396	0.6331	0.0334
REM	0.0524	0.8120	0.0188
Entropy feature evaluation using t-test			
AW	0.102	0.5678	0.0332
S1	0.1002	0.7864	0.1345
S2	0.0345	0.5445	0.0564
S3	0.0456	0.6589	0.0344
REM	0.0112	0.5678	0.0176

3.4 Classifiers

We employed several classification models. In this section, we explain those models.

1. SVM is an efficient model that is used for binary classification. The algorithm employed a hyperplane between two classes to classify samples based on one or more feature samples [17].
2. KNN is un-supervised model that is employed as a clustering. The algorithm finds the number of nearest neighbours among data using a distance metric such as Euclidean distance. Then, it clusters the datapoints based on the results of the first step.
3. Bagging: it is based on a uniform majority voting technique. This technique is used to group the output of classifiers to classify the test sample of an ensemble.
4. Boosting: it's like bagging ensemble. It uses the voting strategy in cases of classification.
5. Stacking**:** it employs a Meta classifier to combine the output of various models in one output. The classification is made by the top layer model using the bottom layer's output.

3.5 Performance evaluation

The following metrics are used for evaluation the proposed model.

- $\text{Accuracy}(\text{ACC}) = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$ (4)
- $\text{Sensitivity}(\text{SEN}) = \frac{\text{TP}}{\text{TP+FN}}$ (5)
- $\text{Specificity}(\text{SPE}) = \frac{\text{TN}}{\text{TN+FP}}$ (6)
- $\text{Precision}(\text{Prec}) = \frac{\text{TP}}{\text{TP+FP}}$ (7)
- $\text{F} - \text{score} = 2 \times \frac{\text{Precision}\times\text{Sensitivity}}{\text{Precision}+\text{Sensitivity}}$ (8)

TP: refers to the total number of stages that are correctly recognized. TN: denotes to the total number of stages who are appropriately classified wrongly. FN: is the total number of stages that were mistakenly classified.

4.Experimental Results

In this paper, a robust model is proposed to classify EEG sleep stages. In this experiment, we discussed the main findings.

To figure out the relationship between sleep stages, and to reflect the actual situation of detection of sleep stages, we classified different classes of sleep-stages. Identifying REM and NREM is essential task for health provider to identify catalepsy disorder. In this paper, we created several classes of sleep stages as described in Table 4.

The classification performance of the proposed model was evaluated based on 10-folds cross-validation metric, specificity and sensitivity and the proposed model was integrated with several machine learning techniques named KNN, Random Forest, Boosted ensemble, Gradient boosting, Bagged Ensemble, SVM, and ensemble algorithms. The EEG data were randomly divided into 10 samples. At each experiment, one sample was used in the testing set and remaining samples were employed as in the training set.

Table 4

Different classes of EEG sleep stages

Classification problem	**Sleep Group**
Six sleep classification issue	AW, N1, N2, N3, N4 and REM
Five Sleep Classification issue	AW, N1, N2, SWS (N3, N4) and REM
Three sleep stages classification issue	AW, NREM, and REM
Two sleep stages classification issue	AW, and sleep (N1-N4, REM)

4.1 Features selection evaluation

First, the extracted features were evaluated to select the most powerful ones. Table 5-7 reports the obtained results based on different combination of features. Table 5 shows the results based on statistical features. As mentioned in section XX, eight statistical features named {*max, range, std, min, mean, median, skewness, kurtosis*} were selected based on feature selection model. The selected features set was used to classify EEG sleep stages. Based on the results, the gradient boosting, stack ensemble, and Boosted ensemble recorded the highest classification average among the models with accuracies of 0.91%, 0.92% and 0.90% respectively. However, KNN ranked the lowest model among classification models.

Another experiment was also conducted to evaluate the linear features. In this experiment, we adopted the two selected features using the feature selection model. The linear features named {*approximation entropy, Shannon entropy}* were used to classify sleep stages. The results showed that the stacked ensemble obtained the best accuracy followed by gradient boosting. However, KNN and Random Forest scored the lowest accuracy among the classification models. The results of six sleep stages are presented in Tables 5 and 6.

The combination of the statistical features and the nonlinear features were tested and used to classify sleep stages. Table 7 reports the sleep classification results. We can observe that the integration of statistical features and the nonlinear features improved the classification accuracy by for all models. it was observed that the stacked ensemble showed a high performance, however, the KNN recorded also the lowest accuracy. We considered the combination of the statistical features and the nonlinear features in our experiment in the next section.

Table 5

Classification results based on statistical features

Classifier	Accuracy	Sensitivity	Specificity	Precision	f-score
SVM	0.7777	0.691	0.7868	0.7632	0.7597
KNN	0.6573	0.6730	0.6650	0.6502	06654
Random forest	0.8194	0.8107	0.8100	0.8203	0.8039
Gradient boosting	0.9114	0.9002	0.9000	0.9110	0.9011
Bagged Ensemble	0.9034	0.9017	0.9102	0.9010	0.9021
Boosted ensemble	0.9115	0.9020	0.9118	0.9033	0.9013
Stacked ensemble	**0.9110**	**0.9121**	**0.9231**	**0.9154**	**0.9186**

Table 6

Classification results based on entropy features

Classifier	Accuracy	Sensitivity	Specificity	Precision	f-score
SVM	0.8035	0.8191	0.8723	0.8131	0.8021
KNN	0.8021	0.8263	0.8123	0.8242	0.8343
Random forest	0.7867	0.7721	0.7854	0.7754	0.8065
Gradient boosting	0.89520	0.8840	0.8922	0.9040	0.8910

Bagged Ensemble	0.8988	0.8939	0.8887	0.8815	0.8923
Boosted ensemble	0.8776	0.8726	0.8641	0.8790	0.8724
Stacked ensemble	**0.9021**	**0.8996**	**0.9110**	**0.9012**	**0.9021**

Table 7

Classification results based on combination of features

Classifier	Accuracy	Sensitivity	Specificity	Precision	f-score
SVM	0.9037	0.8977	0.8985	0.9063	0.9033
KNN	0.8986	0.8876	0.8966	0.8873	0.8852
Random forest	0.9134	0.9144	0.9045	0.9156	0.9183
Gradient boosting	0.9213	0.9180	0.9035	0.9074	0.9152
Bagged Ensemble	0.9424	0.9382	0.9252	0.9283	0.9492
Boosted ensemble	0.9323	0.9371	0.9212	0.9142	0.9353
Stacked ensemble	**0.963**	**0.9733**	**0.9642**	**0.9764**	**0.96531**

4.2 Classification different categories of EEG sleep stages

The performance of proposed model was tested to classify several sleep classification problems. In this experiment, five, two, three sleep classification categories were adopted. The performance of proposed model was reported in Tables 8 and 9. In this experiment, EEG data from two datasets were employed to evaluate the proposed model. we divided the EEG data randomly into 10 groups. At each experiment, one group was used as the testing set, and remaining sets were used as the training sets. The procedure was repeated ten times, and all results were reported. For five sleep classification problem, Table 9 reports the result. We can notice that the proposed model performed well with all classification models. however, the ensemble models outperformed the individual ones. The highest accuracy was obtained by the stacking ensemble with an accuracy 0.96% followed by the Bagged, and Boosted ensembles respectively.

Table 8

Five sleep classification problem

Classifier	**Dataset-1**			**Dataset-1**		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
SVM	0.9132	0.9071	0.8998	0.9012	0.9123	0.9012
KNN	0.8834	0.8832	0.8843	0.8765	0.8743	0.8754
Random forest	0.9031	0.9150	0.9133	0.9100	0.9012	0.9016
Gradient boosting	0.9217	0.9271	0.9242	0.9128	0.9185	0.9122
Bagged Ensemble	0.9343	0.9332	0.9434	0.9321	0.9383	0.9312

Boosted ensemble	0.9331	0.9333	0.9365	0.9234	0.9445	0.9356
Stacked ensemble	**0.9543**	**0.9554**	**0.9531**	**0.9551**	**0.9560**	**0.9617**

With three sleep classification problem, the performance of all models was improved. We record an average of accuracy of 0.96% for ensemble models, and 0.93% for individual models. Table 10 reports the result of three sleep classification problems. We can notice that the proposed model performed was good performance with three-sleep classification problem. The highest accuracy was obtained by stacking ensemble with an accuracy 0.97% followed by the Bagged 0.96%, and Boosted ensembles 096%.

Another problem of sleep stages of two sleep classification problem was discussed in this experiment. From the results, it observed that the performance of all models was improved. The ensemble models recorded accuracies ranging from of 0.98% to 0.98%, and from 0.93% to 0.95% for individual models. Table 11 reports the result of three sleep classification problem. We can notice that the proposed model performed was performed well with all three-sleep classification problems. The highest accuracy was obtained by stacking ensemble with an accuracy 0.98% followed by the Bagged 0.97%, and Boosted ensembles 097%.

Table 10

Three sleep classification problem

Classifier	Dataset-1			Dataset-1		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
SVM	0.9132	0.9071	0.8998	0.9012	0.9123	0.9012
KNN	0.8834	0.8832	0.8843	0.8765	0.8743	0.8754
Random forest	0.9231	0.9256	0.9123	0.9176	0.9167	0.9189
Gradient boosting	0.9267	0.9278	0.9245	0.9123	0.9187	0.9112
Bagged Ensemble	0.9452	0.9532	0.9454	0.9521	0.9583	0.9512
Boosted ensemble	0.9454	0.9543	0.9547	0.9512	0.9521	0.9564
Stacked ensemble	**0.9543**	**0.9587**	**0.9532**	**0.95432**	**0.9576**	**0.9631**

4.3 Performance evaluation based on Confusion matrix

5-fold cross validation was also used for evaluation purpose. The confusion matrices of proposed model using two sleep datasets were reported based on Stacking ensemble in Fig. 5. We found that accuracy of W, and S2 classification were higher than other stages. Our experiment showed that the extracted features of EEG had the ability to recognise stage AW, and stage N2 successfully.

In addition, it was found that some epoch S3 were classified as S2. The reason was that both stage S2 and stage S3 had some similar statistical features which made them difficult to identify. This was appeared in Datset_1 dataset.

True Label	W	S1	S2	S3	REM
W	0.97	0.00	0.01	0.01	0.01
S1	0.011	0.95	0.01	0.01	0.01
S2	0.01	0.009	0.95	0.009	0.018
S3	0.003	0.01	0.01	0.94	0.001
REM	0.001	0.001	0.001	0.02	0.95

Predicted label

True Label	W	S1	S2	S3	REM
W	0.96	0.01	0.03	0.02	0.00
S1	0.03	0.95	0.01	0.001	0.00
S2	0.01	0.001	0.95	0.01	0.02
S3	0.01	0.04	0.05	0.90	0.000
REM	0.001	0.001	0.01	0.02	0.96

Predicted label

Fig 4. confusion matrix of the proposed model

4.4 Comparative with EEG sleep stages models

To evaluate the proposed model in EEG sleep stages classification, we conduced comparisons with some existing sleep classification models. The output of comparisons was given in Table 12. We considered some metrics in the comparisons including accuracy and sensitivity. In addition, five and six sleep stages classification problems were included in these comparisons. We can notice that the proposed model performed better in six and five sleep stages problems than other models. The results of comparisons stated that the proposed model can contribute and add a significant impact EEG sleep stages research. It can be used as hardware system to by experts for identifying of sleep disorders.

Table 12

Comparisons among the proposed model with previous methods

Authors	Model	Channel used	Accuracy
Doroshenkov et al., [36]	HMM	Pz-OZ, Fpz-Cz,	62%
Ebrahimi et al., [37]	Wavelet based on ANN	, Pz-OZ, Fpz-Cz	-
Hassan et al., [38]	TQWT model with RF	Pz-Oz	93.37%
Hassan et al., [39]	TQWT based on ensemble	Pz-Oz	92.44%
Hsu et al., [40]	Energy features based Model	Pz-OZ, Fpz-Cz,	-
Liang et al., [41]	AR model based on LDA	Fpz-Cz	76.72%
Zhu et al., [42]	Visibility graph	Fpz-Cz	87.51%, 0.82%
Berthomier et al., [43]	Fuzzy logic approach	Fpz-Cz	-
Ronzhina et al., [45]	Power spectral density model with LDA	Pz-OZ	76.71%
Abdulla et al., [46]	Correlation graph-based model	Pz-OZ	93%
Diykh et al., [47]	undirected graph Model	C3-A, Pz-OZ,	95.5%
Diykh et al., [48]	Graph model	Pz-OZ	92.1%
The proposed model	**DWT based linear and entropy features**	**Pz-OZ**	**96%, 97%**

5.Conclusion

In this paper, we proposed an intelligent model for sleep stages classification. DWT was utilised to analyse EEG signals. Each EEG segment was decomposed by DWT and g statistical and nonlinear features were extracted. These features were tested using statistical metrics. The results proved that the proposed model could improve the performance of sleep classification making small step toward application in actual situation.

References

[1] **Goldberger, A. L., et al. (2000).** PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation, 101(23), e215–e220. doi: 10.1161/01.CIR.101.23.e215.

[2] **Shao, Y., et al. (2024).** Reliable automatic sleep stage classification based on hybrid intelligence. Computers in Biology and Medicine, 173, 108314. doi: 10.1016/j.compbiomed.2024.108314.

[3] **Cheng, X., Huang, K., Zou, Y., & Ma, S. (2024).** SleepEGAN: A GAN-enhanced ensemble deep learning model for imbalanced classification of sleep stages. Biomedical Signal Processing and Control, 92, 106020. doi: 10.1016/j.bspc.2023.106020.

[4] **Abdulla, S., Diykh, M., Siuly, S., & Ali, M. (2023).** An intelligent model involving multi-channels spectrum patterns based features for automatic sleep stage classification. International Journal of Medical Informatics, 171, 105001. doi: 10.1016/j.ijmedinf.2023.105001.

[5] **Gao, V., Turek, F., & Vitaterna, M. (2016).** Multiple classifier systems for automatic sleep scoring in mice. Journal of Neuroscience Methods, 264, 33–39. doi: 10.1016/j.jneumeth.2016.02.005.

[6] **Zhu, G., Li, Y., & Wen, P. (2014).** Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. IEEE Journal of Biomedical and Health Informatics, 18(6), 1813–1821. doi: 10.1109/JBHI.2014.2306423.

[7] **Shi, J., Liu, X., Li, Y., Zhang, Q., Li, Y., & Ying, S. (2015).** Multi-channel EEG-based sleep stage classification with joint collaborative representation and multiple kernel learning. Journal of Neuroscience Methods, 254, 94–101. doi: 10.1016/j.jneumeth.2015.07.003.

[8] **Kayikcioglu, T., Maleki, M., & Eroglu, K. (2015).** Fast and accurate PLS-based classification of EEG sleep using single channel data. Expert Systems with Applications, 42(21), 7825–7830. doi: 10.1016/j.eswa.2015.06.009.

[9] **Huang, C.-S., Lin, C.-L., Ko, L.-W., Liu, S.-Y., Su, T.-P., & Lin, C.-T. (2014).** Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels. Frontiers in Neuroscience, 8, 263. doi: 10.3389/fnins.2014.00263.

[10] **Rodríguez-Sotelo, J. L., Osorio-Forero, A., Jiménez-Rodríguez, A., Cuesta-Frau, D., Cirugeda-Roldán, E., & Peluffo, D. (2014).** Automatic sleep stages classification using EEG entropy features and unsupervised pattern analysis techniques. Entropy, 16(12), 6573–6589. doi: 10.3390/e16126573.

[11] **Samiee, K., Kovács, P., Kiranyaz, S., Gabbouj, M., & Saramäki, T. (2015).** Sleep stage

classification using sparse rational decomposition of single channel EEG records. In 2015 23rd European Signal Processing Conference (EUSIPCO) (pp. 1860–1864). IEEE. doi: 10.1109/EUSIPCO.2015.7362747.

[12] **Aboalayon, K. A. I., Ocbagabir, H. T., & Faezipour, M. (2014).** Efficient sleep stage classification based on EEG signals. In IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014 (pp. 1–6). IEEE. doi: 10.1109/LISAT.2014.6845196.

[13] **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. doi: 10.1109/TNSRE.2016.2517824.

[14] **Doroshenkov, L. G., Konyshev, V. A., & Selishchev, S. V. (2007).** Classification of human sleep stages based on EEG processing using hidden Markov models. Biomedical Engineering, 41(1), 25–28. doi: 10.1007/s10527-007-0006-5.

[15] **Ebrahimi, F., Mikaeili, M., Estrada, E., & Nazeran, H. (2008).** Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 1151–1154). IEEE. doi: 10.1109/IEMBS.2008.4649313.

[16] **Hassan, A. R., & Bhuiyan, M. I. H. (2016).** A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. Journal of Neuroscience Methods, 271, 107–118. doi: 10.1016/j.jneumeth.2016.07.001.

[17] **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. doi: 10.1016/j.knosys.2017.05.014.

[18] **Hsu, Y.-L., Yang, Y.-T., Wang, J.-S., & Hsu, C.-Y. (2013).** Automatic sleep stage recurrent neural classifier using energy features of EEG signals. Neurocomputing, 104, 105–114. doi: 10.1016/j.neucom.2012.10.019.

[19] **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(6), 1649–1657. doi: 10.1109/TIM.2012.2187210.

[20] **Berthomier, C., et al. (2007).** Automatic analysis of single-channel sleep EEG: validation in healthy individuals. Sleep, 30(11), 1587–1595. doi: 10.1093/sleep/30.11.1587.

[21] **Ronzhina, M., Janoušek, O., Kolářová, J., Nováková, M., Honzík, P., & Provazník, I. (2012).** Sleep scoring using artificial neural networks. Sleep Medicine Reviews, 16(3), 251–263. doi: 10.1016/j.smrv.2011.06.003.

[22] **Abdulla, S., Diykh, M., Laft, R. L., Saleh, K., & Deo, R. C. (2019).** Sleep EEG signal analysis based on correlation graph similarity coupled with an ensemble extreme machine learning algorithm. Expert Systems with Applications, 138, 112790.

[23] **M. Diykh, Y. Li, and S. Abdulla,** "EEG sleep stages identification based on weighted undirected complex networks," Computer Methods and Programs in Biomedicine, vol. 184, p. 105116, 2020. https://doi.org/10.1016/j.cmpb.2019.105116