Suren Kumar Sahu[1*], Sudhir Kumar Mohapatra[2*], Santosh Kumar Satapathy[3*]

Multi-Channel EEG Analysis for Automated Sleep Staging Characterization Using Ensemble Learning Techniques

Journal of Electrical Systems

Abstract. Advancements in electrical engineering and signal processing can enable the development of automated sleep data analysis systems that embody significantly more advantages than their traditional manual counterparts. Recent approaches to automated sleep stage classification are considered from the PICO framework (Population, Intervention, Comparison, Outcome). Traditionally, PSG/EEG-based sleep studies were a laborious and time-consuming process. Despite this, significant promise is being offered from promising stride development in machine learning (ML) and deep learning (DL) technologies for full automation of sleep stage scoring. The study is applied to data from the SleepEDF public PSG Hypnogram dataset, which combines EOG signal processing and other physiological signals to classify sleep into five stages. The raw signals are segmented into epochs and passed through four machine learning models: Random Forest, Gradient Boosting, Bagging Classifier, and Ensemble Learning. Those accuracies were classified using criteria such as accuracy percentages of 78% for Random Forest, 79% for Gradient Boosting, 75% for Bagging Classifier, and 85% for Ensemble Learning. The comparison has demonstrated Ensemble Learning as the most accurate model and, hence, has shown prospects for being implemented into consumer-grade sleep monitoring devices. Besides highlighting the advantages of automated systems in classification tasks of sleep stages, the results of the experiment suggest possible applications in real-world sleep tracking.

Keywords: Automated sleep stage scoring (ASSS), Machine Learning (ML), Electroencephalogram (EEG), Feature Engineering (FE), Electrooculogram (EOG).

1 INTRODUCTION

Sleep forms a core part of human health, influencing different physiological activities undertaken in carrying out daily chores, mental well-being, and overall health [1]. Lack of disturbance of sleep may result in restless legs syndrome, sleep apnea, insomnia, and other related disorders which have become very common due to the modern lifestyle of people that involves much stress. All these sleep-related disorders can be controlled through the reading of the bioelectric signals, called the electroencephalogram (EEG). Over the past few years, enhancement in electrical and signal processing has led to the execution of automated sleep data analysis Traditionally, techniques like polysomnography and EEG recordings are rather time-consuming.[3] However, progress with machine learning and deep learning shows promise in automating these processes [2]. The discussion will base itself on the benefits versus challenges of such modern techniques, which might become part of consumer devices shortly. In this work, we will discuss how wearable technology could be used for extracting robust features, and how machine learning algorithms can be used to enhance accuracy close to that of medical professionals. Despite improvements in sleep scoring methods, huge gaps exist. Sleep is a state of decreased activity—both mental and physical—in which consciousness is altered and several sensory activities are temporarily suspended. Less interaction is characterized by sleep with the environment; the period is marked by wakefulness and loss of consciousness. Most important processes, however, do occur during sleep, such as dreaming, changes in brain activity, neuron regeneration, muscle clearance, and in other aspects, overall health maintenance. Sleep itself reduces the risk of chronic illnesses, lessens stress, improves the health of the mind, and even reduces the risks for the most severe illnesses, which are type 2 diabetes and heart diseases. Sleep further maintains a functioning metabolism and enhances immunity, therefore being able to actively fight germs for the constant health of the body.[5] In younger people, sleep is important as it increases growth and development and improves memory and learning functions. Bio signals are those physiological signs emitted by biological beings.

A Biosignal could be a biological event over time or naturally: for example, breathing, heartbeats, and signal transmissions in the brain. We are performing diverse techniques to capture such kinds of events when they occur. For example, we record brain waves using an electroencephalograph (EEG) and heartbeats using an electrocardiogram (ECG). Whenever we record the combination of multiple signals, then it is called polysomnography (PSG). Other kinds of biosignal with waveforms reflecting on various physiological events include electromyogram (EMG), electrogastrogram (EGG), electroretinogram (ERG), electrooculogram (EOG), phonocardiogram (PCG), and galvanic skin. II. BRAIN WAVE CHARACTERISTICS IN SLEEP STAGES AASM has published a paper in which they investigated some data extraction techniques with the help of the following domains: 1. Time 2. Frequency 3. Time and Frequency 4. Nonlinear and Entropy Domain As we can see nowadays signal processing, statical analysis, and Computer Science are getting revolutionized however there is a gap between machine learning and sleep staging automation There are some barriers to the adoption by sleep data professionals we need to collaborate between academia, research, and industry which can handle the large dataset for the sleep and signal processing and can manage these methods For analysis purposes, we are using 3 commonly used signals in sleep studies: 1. Electroencephalogram (EEG) 2. Electrooculogram (EOG) 3. Electromyogram (EMG)

*Faculty of Emerging Technology, Sri Sri University, Cuttack, Odisha, India. [1]suren.sahu2602@gmail.com
*Faculty of Emerging Technology, Sri Sri University, Cuttack, Odisha, India [2]sudhir.mohapatra@srisriuniversity.edu.in
*Information and Communication Technology, Pandit Deendayal Energy University,382007, Gandhinagar, Gujarat, India.[3]santosh.satapathy@sot.pdpu.ac.in.

The sleep cycle consists of 3 stages that include nonrapid eye movement (NREM) and rapid eye movement (REM) [2].Fig.1 illustrates the representation of the sleep rhythms captured during the sleep hours.

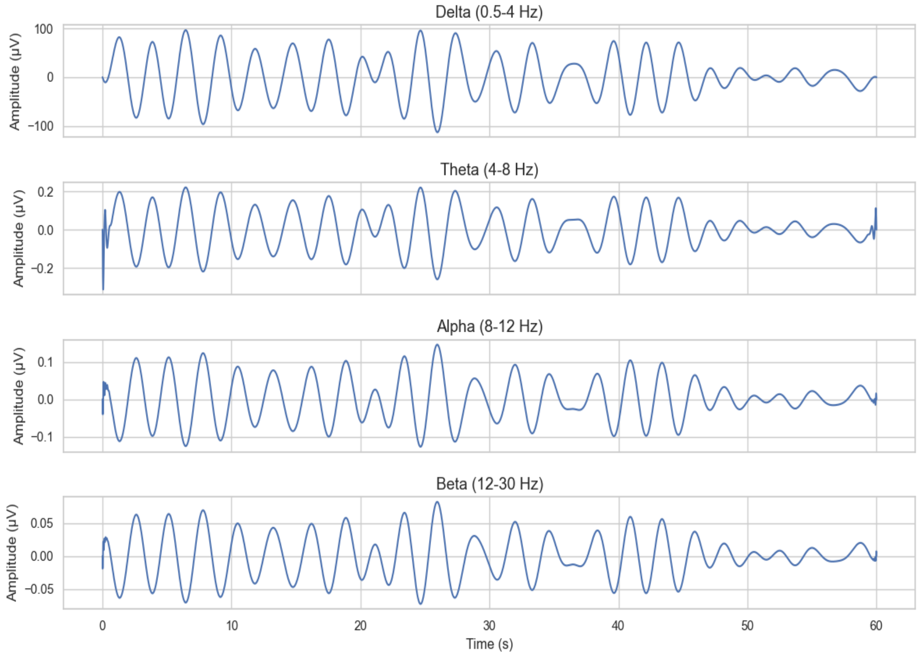

Fig. 1.Different brain EEG rhythms reported during sleep hours

Some of the important signals used in sleep studies There are three signals used for sleep staging: Electroencephalogram (EEG) measures electrical activity in the brain and gives information about different sleep stages based on the presence of specific wave frequencies. Electrooculogram (EOG) monitors eye movements. It becomes useful in ascertaining REM sleep, which is identified by the occurrence of rapid eye movements. Electromyogram (EMG)This is a measure of muscle activity and helps in distinguishing between different stages of sleep. It is especially useful in identifying the muscle atonia of REM sleep. Sleep Cycle and Stages The sleep cycle includes different stages, notably nonrapid eye movement, and rapid eye movement sleep. NREM further breaks down into three stages: Stage 1 (NREM1): Transition from wake to light sleep. Theta waves can be recorded in this stage with frequencies ranging from 4 Hz to 8 Hz. Stage 2 (NREM2): Deep sleep. The K-complexes are 0.5 to 2 Hz in frequency, while the sleep spindles are. Stage 3 (NREM3): Deep sleep, otherwise referred to as slow wave sleep, is characterized by delta waves with frequencies within the range of 0.5 to 4 Hz.[7] EEG Data and Preprocessing Most EEG data is stored in the European Data Format and typically recorded from multiple channels. The data is generally captured to retain the spatial information that comes from different parts of the brain.

Regularization: Quality of the signal, consistency-related techniques. Feature Extraction: Features relevant for analysis and classification are extracted from the pre-processed data. Data Labelling: According to the American Academy of Sleep Medicine, the data is labeled to standardize the sleep stage classification methods. Feature Extraction and Machine Learning Feature extraction is, therefore, a very important process or aspect of EEG data analysis for sleep staging.
It shall be applied to find and quantify those features of the signal indicative of the different stages of sleep. Common features include Power Spectral Density [3] Which measures the power of the signal across different frequency bands. Amplitude and Frequency of Brain Waves: It identifies the dominant brain wave types present in the signal. Entropy Measures: It provides a metric about the complexity and variability of the signal. Extracted features are used along with machine learning

2 RELATED WORK

This paper is the perspective of what physiological arousal during sleep does, especially on aspects of impairments in cognition changes and heart rate variations. The interruption to the continuity of sleep brought by arousals impacts a person's overall quality of rest due to impaired cognition function and unstable heart rates. For this, the researchers dealt with the problem using EEG signals and machine learning algorithms. It correctly identified arousal patterns with its model and with an intensity level as well as sensitivity at 82.68%, and specificity at 95.68% with an AUROC (Area Under the Receiver Operating Characteristic curve) of 96.30%. It does seem to be potent enough for qualifying physiological disruptions during sleep in various stages [1].
In this paper new methods for time and frequency domains for dealing with mis-classified data using ensemble learning strategies. It is a multi-model machine learning approach toward class inaccuracies and ensures better reliability by having increased model dependency. Feature extraction has been carried out based on guidelines of the American Academy of

Sleep Medicine for sleep stage identification [9]. Class-balanced random sampling techniques were used to help in the improvement of the model with enhanced high accuracy as well as high mean F1 scores. This approach showed high robustness for classifying sleep stages in individuals and provided scope for ensemble learning in improving the analysis of the sleep stages [2]. A particular article about sleep stage classification used polysomnographic signals to classify them. These included EEG, EMG, EOG, and ECG signals. They attempted to capture the interdependency and the nonlinear nature of complexity that lay within the signals to be able to better distinguish between the states of sleep. A model was developed using a small dataset, which established that it could be able to discriminate between different sleep states with 74% accuracy. Multiple bioelectrical signals become essential to interpret this sleep pattern if there's not enough data to understand [3].

Another study revolves around the automation of polysomnography, which traditionally has been a completely manual procedure, with deep learning models-RCNN. The same model was trained to automate the classification of sleep stages and apneas/sleep, as well as limb activities otherwise done by human experts. The outcome of the result achieved a sleep staging with an accuracy of 87.6%, 88.2% for the detection of sleep apnea, and 84.7% for the detection of limb movements. These results indicate how the capabilities of RCNNs can be used to mimic human intervention and facilitate fast, efficient, and accurate sleep analysis with other tasks without losing near-perfect accuracy [4].

In this paper technological advancement and automation in sleep stage scoring, which has been transformed through adaptive and parallel computing models such as artificial neural networks. One can note that these models assure promising solutions for fast, efficient, and reliable classification of a sleep stage. Their application allows the processing of vast datasets with high speeds but yields no deterioration in performance due to the use of ANN. This would be the future of sleep analysis where automation and computational power work hand-in-hand with clinicians to get rapid and accurate insights into a patient's sleep pattern [5]. A third study utilized EEG data for sleep stage classification by using feature extraction based on standardized AASM guidelines. Researchers utilized ensemble learning in sleep stage classification using class-balanced and random sampling techniques. Measurement of the outcomes included mean F1 scores and accuracies for each sleep stage; the method demonstrated strong and consistent performance. In the article, the adoption of standardized guidelines for extracting features in sleep research was highlighted, hence obtaining consistency and clinical relevance of the results [6].

In another article, researchers compared machine learning techniques, especially feature extraction from bioelectric signals with traditional methodologies. The intention was to improve the identification of the sleep stage, closing the gap between the state-of-the-art machine learning models and established signal processing techniques. The quality of the study was evaluated using the PICO framework, and it mainly focused on the need for robust models that are clinically useful in the automation of sleep staging. The researchers cited the shortcomings of existing methodologies and argued the need to establish clinically applicable yet accurate machine-learning models. The work ends by stating that although techniques adopted for the development of machine learning algorithms hold much promise for automating sleep stage classification, in clinical practice daily, it has been sorely lacking [6]. Their research states how both automated and semi-automated deep learning techniques can revolutionize sleep research. However, the study concluded that more diverse datasets with accurate development are needed to achieve reliable and notable results for the application of the models in clinical practice. Researchers observed that datasets from diverse sources are important in training the generalization of models across different populations and sleep conditions [7].

2.1 Contribution

This study will extend the existing work of research on this topic as it will employ several machine learning models and EEG-based feature extraction techniques for sleep-stage classification through the utilization of a Random Forest Classifier to validate the performance of the model. The analysis employed EEG, EMG, and EOG signals from the Sleep Physionet dataset and focused on both deep learning and ensemble techniques to increase the accuracy rates. This code has been specially designed for the processing of raw sleep recordings, feature extraction, resampling of data using SMOTE, and its classification using a Random Forest model for the stages of sleep with physiological signals. This work significantly contributes toward the automation of the sleep stage classification by exploiting advanced signal processing techniques and machine learning models, like Random Forests, when dealing with the complexity and nonlinear nature of sleep data.

This research begins with loading and preprocessing sleep data in EDF format that filters EEG channels and sets up annotations that represent sleep stages. Then, the code applies a bandpass to limit the frequency range for better feature extraction, and epochs representing sleep stages are created from filtered signals. This way, the data is divided into time intervals appropriately and processed with la-bels during the sleep stages.

The current study also addresses imbalanced datasets in sleep research. Imbalance is a very common feature of sleep data because some of the sleep stages, for instance, have fewer records than others. To prevent this, the dataset was made balanced with the application of SMOTE (Synthetic Minority Over-sampling Technique), which significantly enhanced the capabilities of the model on the classification of minority classes. In this sense, the model will not overfit the more frequent stages such as light sleep and wakefulness.

The use of the Random Forest Classifier is also one of the contributions of the study. By using this ensemble learning approach, the research therefore gains improved performance of classification because it is very resilient to handling high dimensional data which usually takes place in polysomnography. The resampled dataset is used for training the classifier to then test performance using confusion matrices and classification reports. The study reported a high classification

accuracy across the different stages of sleep and further describes this in a confusion matrix, where one could illustrate the ability of the model to differentiate between wakefulness, light sleep-N1, N2, deep sleep-N3, and REM.
The study also presents a new contribution in the form of a heat map visualization of the classification performance; in this case, the confusion matrix describes the performance of a Random Forest model. Using this can potentially help the model to identify where it's doing well and where it is not performing so well thereby pointing to possible directions for further improvements. Feature extraction in EEG data with SMOTE resampling combined with ensemble techniques such as Random Forest contributes significantly towards automation in the classification of sleep stages. Dealing with technical difficulties as well as clinical challenges related to sleep research are also involved in the current study.
It is with this backdrop that the authors have made tremendous contributions to automation in sleep staging through open-access data, and more advanced machine learning techniques, and addressed issues such as class imbalance, which often dominate a lot of problems in sleep staging. Thus, its approach presents a scalable but robust alternative that eventually can be used clinically for automated sleep analysis.

3 METHODOLOGY

We now used the raw EEG data of Sleep Physionet to automate this process in this research study using machine learning techniques. We applied filtering on the EEG channels and created epochs for segmenting the periods before we classified the sleep stages. We employed synthetic oversampling using the SMOTE algorithm to solve the class imbalance problem. We finally applied the resampled data for training the Random Forest Classifier to predict sleep stages. Finally, we classified and presented the performance of the model using the classification report and confusion matrix, where the model may be applied to detect the presence of multiple sleep stages with excellent accuracy.Fig.2 presents complete architecture of the proposed research work.

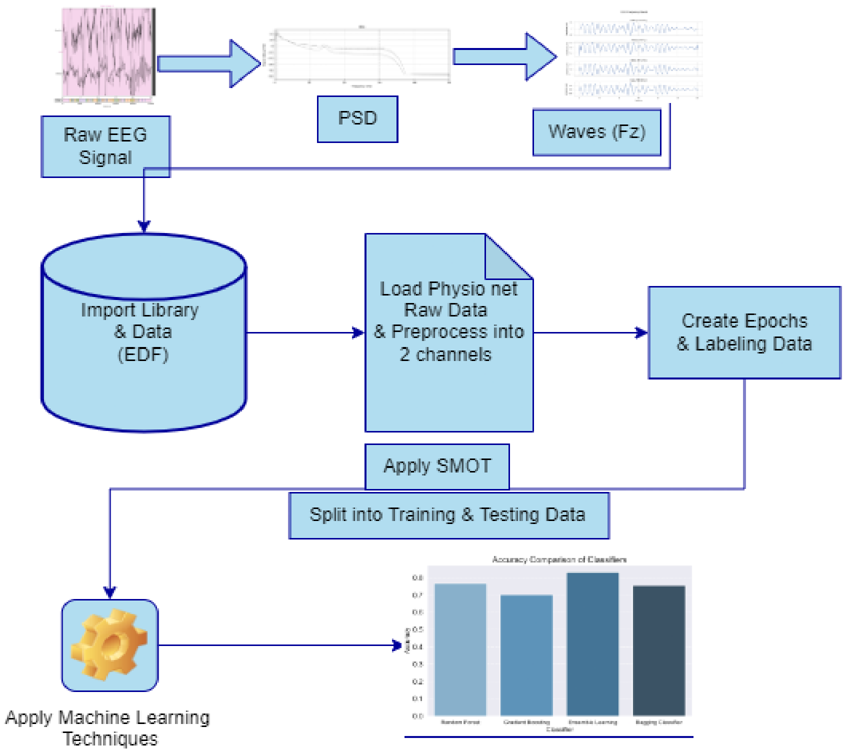

Fig. 2. An overview architecture of a multi-channel automated sleep staging system

3.1 Dataset Acquisition

The data preparation face involves loading and preprocessing of the EEG data the raw EDF data are obtained from several EDF (European data format) data files we have used the MNE library in Python to load MNE data EDF files are loaded from the local machine we specify the path and collect the EDF hypnogram from that for analyses the EDF data we have a specific function load sleep physio net raw which is defined for the specific load and preprocessing the data Its use to load EDF data as an integral part of data preprocess pipelines is used to create epochs and classification of data further into it So let’s discuss the function parameters that we are passing and how it works we have raw names, and annotate filename, load EEG signal, crop wake mins so raw FNAME contains the file data which contains EDF data so annotate filename indicates path name which provide event marker corresponding to sleep stages 3rd parameter was load EEG signal only which exclude the rest data which is not EEG signal so we can exclude further signal which is not required crop wake mins is useful to remove the wake state of the signal in which we remove the signal from starting and ending which include the wake state which not helpful for sleep analysis. The recorded sample sleep EEG data is presented in Fig.3.

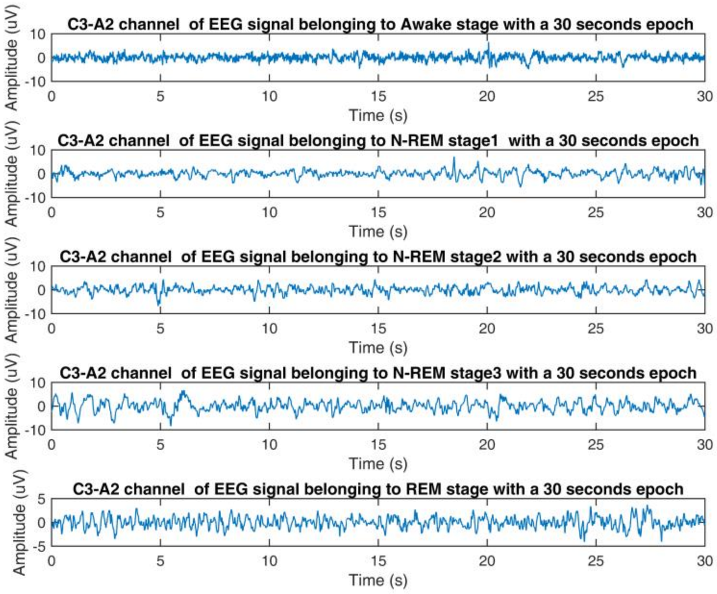

Fig. 3. Illustrations of individual sleep stages behavior

So, our first step is channel map-ping and exclusion the function starts by defining a mapping for non-EEG channels which include horizontal EOG, respiratory signals, submental EMG, and other signals if load EEG is only True then the rest of the signals will be removed as we won't require them secondly we load raw and annotations which will be done by MNE library function default functions than we do is crop wake mins parameter is greater than zero it crop first and last mins wake stages its Boolean filtration by as explained above that range we will filter out the rest data than we will be renaming the EEG channel and mapping into their new names with than the base file is extracted for gather the subject and recording information the data stored into the raw dictionary this its basic data structure which will help us to easy access the important metadata and subsequent analysis and then we return raw which is pre-processed. Then we load our EEG dataset into memory for further analysis and visualizing the EEG signal helps us to inspect data quality and characteristics then we pass high pass and low pass filters to remove outliers we applied 30 Hz purpose of it is to remove high pass frequency noise that can obscure the underlying EEG Signals

The next section is to create Epochs creation and classifying the data so first we create epochs of 30 second which is common duration for analyzing sleep stages than we generating epochs and storing and labelling data corresponding the stages they are and annotate into sample frequency than we do even filtering this will we give each stage their own ID's according to its stage names for our purpose of study we required 5 stages Wake , N1 , N2 , N3 and REM [1] we filter this event and according to its even we classify further into their event stage names this filtering process will make ensure that only sleep events will go for further process rest signals which is not classify into this stages will remove so its our end of preprocess data stage after further we will make sure that rest of data will properly store into NUMPY array for further process and then we will make validation steps for further dimension checks to see whether our array match the size of validation with filter event dimension and filter event shape matching .

After creation of epochs we labelling the epochs and return the epochs and la-bels once its done we do data aggregation and reshaping if it's not get match than data is ready for classification task the major issue occurred is class over-sampling or under sampling where one minor class get less weightage than other class for solving this issue we have sued SMOT techniques which will know as synthetic minority over sampling technique which is popular machine learning technique which will increase the minority class sampling with computable to majority class.

3.2 Data Preprocessing

The dataset used in this paper is based on EEG signals sourced from the Sleep Physio net dataset. Having obtained the raw EEG signals, their hypnogram annotations have also been considered for reference. Recordings are taken for 30 subjects, and data for each subject is recorded in EDF files. The following steps were realized for data acquisition and preprocessing The code collected all the EDF files in a given directory, './path. Their pathways had been accessed programmatically for the purposes of loading. Data Loading: Each recording of EEG has been loaded using the library MNE, designed to process neurophysiological data. The raw data consisted of EEG channels; any non-EEG channels were excluded, such as EOG or EMG, to facilitate analysis. Annotation Handling. Reading the corresponding annotations to identify the stages of sleep.

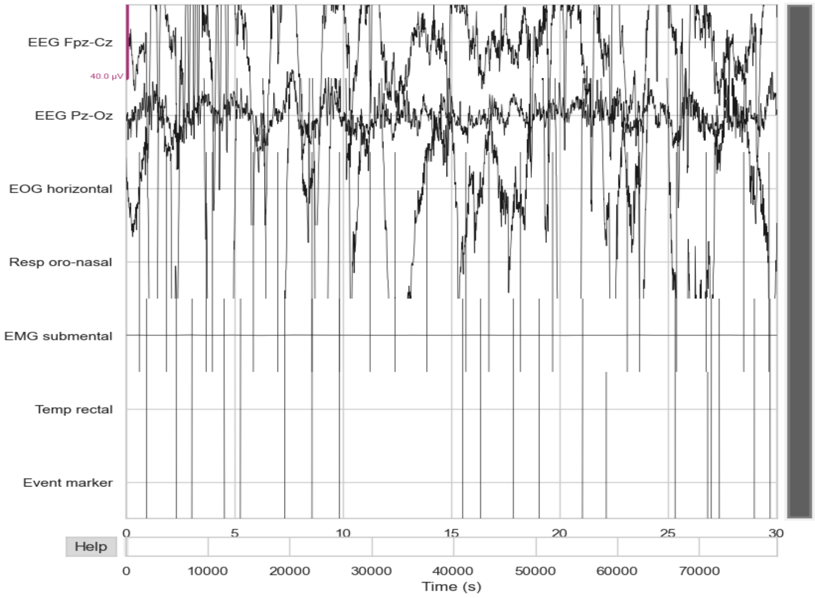

Fig. 4. Signals captured before Preprocess Channels

The dataset used in this paper is based on EEG signals sourced from the Sleep Physio net dataset. Having obtained the raw EEG signals, their hypnogram annotations have also been considered for reference. Recordings are taken for 30 subjects, and data for each subject is recorded in EDF files. The following steps were realized for data acquisition and preprocessing The code collected all the EDF files in a given directory, './path. Their pathways had been accessed programmatically for loading. Data Loading: Each recording of EEG has been loaded using the library MNE, designed to process neurophysiological data. The raw data consisted of EEG channels; and any non-EEG channels, such as EOG or EMG, to facilitate analysis. Annotation Handling. Reading the corresponding annotations to identify the stages of sleep. Only sleep-related events were retained, from stages 1 (N1) through stage 5 (REM), and the wake events were filtered to keep only the stages of sleep. Data Cropping: The data were cropped for wake periods before and after the sleeping events, considering a specified duration of 30 minutes to ensure that just sleeping data were considered for analysis. Creating Epochs: To have a better analysis of the EEG signals, the continuous data was segmented into epochs. Epoch Ing: The continuous raw recording was divided into 30-second epochs. The MNE library's Epoch Ing functions were used along with filtered events corresponding to the identified stages. Labeling: Each epoch was labeled with its corresponding sleep stage from the annotations to make it a complete dataset for supervised learning.

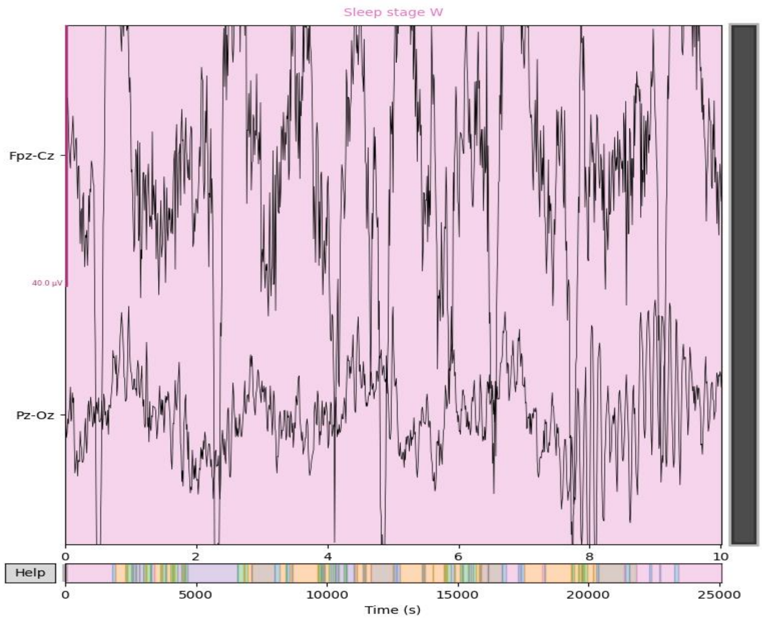

Fig. 5. Signals captured after Preprocess Channels

3.3 Data Reshaping and Balancing

These concatenated epochs were reshaped to be fed into machine learning models. The data were flattened from a three-dimensional shape to two dimensions with the structure of the number of epochs × several channels × several time points, making it suitable for training classifiers. Class Distribution Check: The distribution of the labels was checked for any class imbalances. Some sleep stages were underrepresented. SMOTE was used to generate synthetic examples of the minority class to balance their representation in the dataset for training, in situations of class imbalance.

3.4 Feature Extraction

In signal processing, the Power Spectral Density (PSD) quantifies how the power of a signal or time series is distributed across different frequencies. For a real-valued signal x(t)x(t)x(t), the PSD provides insights into the dominant frequency components, which is particularly useful in analyzing signals from physiological systems such as EEG or ECG. PSD is computed as the squared magnitude of the Fourier Transform, normalized by the time window over which the signal is observed.

In practical applications, PSD can be estimated using methods such as Welch's method, which divides the signal into overlapping segments and averages the periodogram of each segment to reduce variance. For instance, in this study, we employ PSD analysis to understand the spectral characteristics of EEG signals during different sleep stages.

This equation and explanation can be adapted or extended based on the focus and scope of your research.

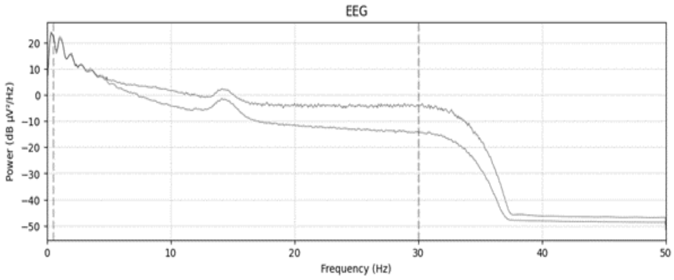

Fig. 6. Representation of power spectral density feature

Then we applied power spectral density power spectral density helped us to understand how the power of a signal or time series is distributed with frequency it helps us to get insightful information from the signal. Power spectral density (PSD) is a measure of how a signal's power is distributed across different frequencies. It's also known as the power spectrum.

3.5 Classification Algorithm

Random Forest This algorithm is a powerful tree-learning technique in Machine Learning. It works by creating several Decision Trees during the training phase. Each tree is constructed using a random subset of the data set to measure a random subset of features in each partition. This randomness introduces variability among individual trees, reducing the risk of overfitting and improving overall prediction performance.Fig.7 illustrates the representation of RF representation.

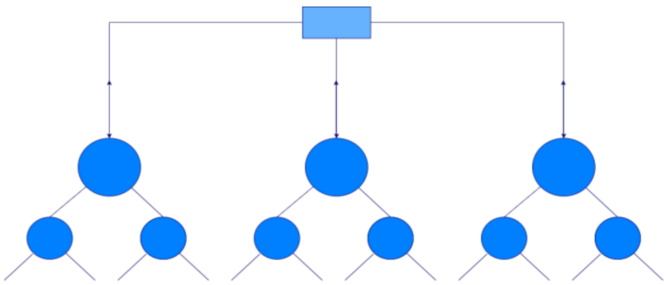

Fig. 7. Random Forest

Gradient Boosting This algorithm is a powerful boosting algorithm that combines several weak learners into strong learners, in which each new model is trained to minimize the loss function such as mean squared error or cross-entropy of the previous model using gradient descent. In each iteration, the algorithm computes the gradient of the loss function concerning the predictions of the current ensemble and then trains a new weak model to minimize this gradient. The predictions of the new model are then added to the ensemble, and the process is repeated until a stopping criterion is met.Fig.8 presents the sample work structure of the gradient boosting.

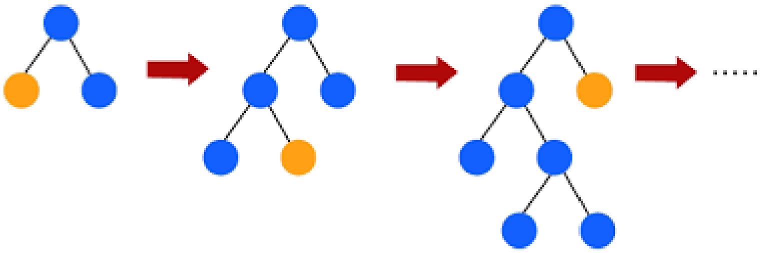

Fig. 8. Gradient Boosting

Ensemble Learning. This is a machine learning technique that combines the predictions from multiple individual models to obtain a better predictive performance than any single model. The basic idea behind ensemble learning is to leverage the wisdom of the crowd by aggregating the predictions of multiple models, each of which may have its strengths and weaknesses. This can lead to improved performance and generalization.Fig.9 shows the working principles of ensemble learning.

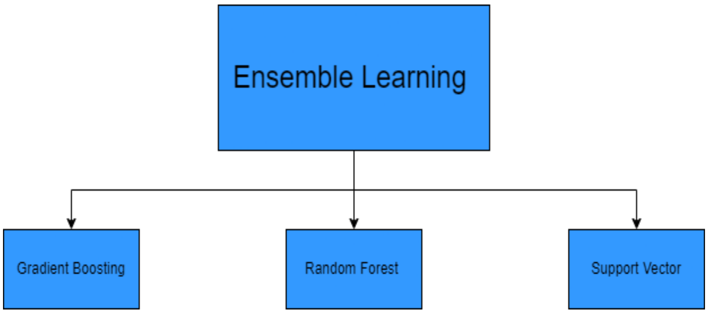

Fig. 9. Ensemble Learning

Bagging Classifier Bagging (or Bootstrap aggregating) is a type of ensemble learning in which multiple base models are trained independently and in parallel on different subsets of the training data. Each subset is generated using bootstrap sampling, in which data points are picked at random with replacement. In the case of the bagging classifier, the final prediction is made by aggregating the predictions of the all-base model using majority voting. In the models of regression, the final prediction is made by averaging the predictions of the all-base model, and that is known as bagging regression.Fig.10 presents the working principles of the bagging classifier.

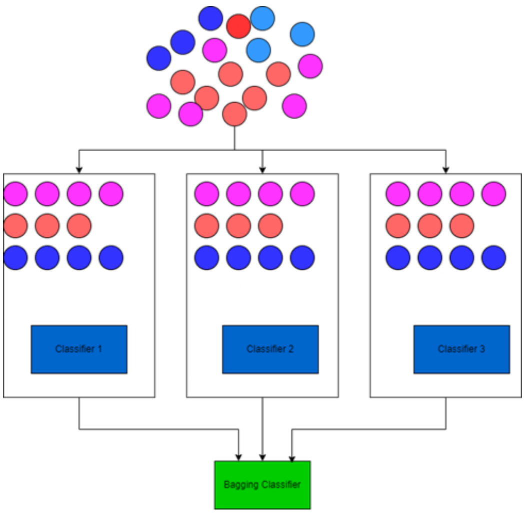

Fig. 10. Bagging Classifier

3.6 Model Training and Testing

A random forest classifier was chosen for training and evaluation due to its robustness in handling a variety of feature types and its capability to manage high-dimensional data: Data Splitting: The resampled data was split into an 80-20 ratio for training and testing sets, respectively, to validate model performance. Model Training: The Random Forest Classifier was initialized and fitted with the training dataset. Hyperparameters were set to use 100 estimators, enhancing its generalization capability. Prediction and Evaluation: The model was used to make predictions on the test dataset, and performance evaluation was done using classification metrics such as accuracy, precision, recall, and F1-score. Confusion Matrix Visualization: A confusion matrix was generated for a graphical representation of the classifier's performance across different sleep classes, displayed using Seaborn's heatmap functionality to provide insights into misclassifications.

3.7 Performance Metrics

Accuracy

Accuracy measures the overall proportion of correctly predicted instances (both true positives and true negatives) out of the total number of instances. It is defined as:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision

Precision calculates the proportion of correctly predicted positive instances out of all instances predicted as positive. It is given by:

Precision = TP / (TP + FP)

Recall

Recall measures the proportion of actual positive instances that the classifier correctly identified. It is expressed as:

Recall = TP / (TP + FN)

F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balanced metric that accounts for both false positives and false negatives. It is computed as:

F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

4 RESULT ANALYSIS AND DISCUSSION

4.1 Random Forest

In this study, we used several machine learning techniques to classify our sleep stage classification first technique is the random forest classifier which is a widely adopted ensemble learning it’s a very versatile method for several problem statements and also it gives high accuracy when working with structured tabular data by constructing multiple decision trees random forest reduce the risk of overfitting and enhance generalizability in this model training set up we are breaking the training set into 80 % and testing on 20 % rest data, 42 random states the larger number of trees can increase accuracy but also it should be maintained computation so we use 100 trees in our model for the classification purpose we use f1 score, precision score, recall score, and accuracy these numbers show how good our model for distinguish the classes. Fig.11 presents the confusion matrix for the RF classification model.

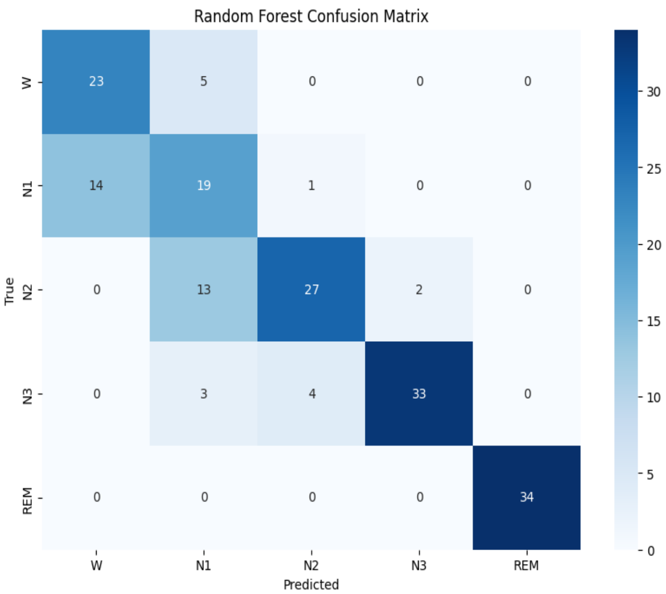

Fig. 11. Confusion Matrix Random Forest

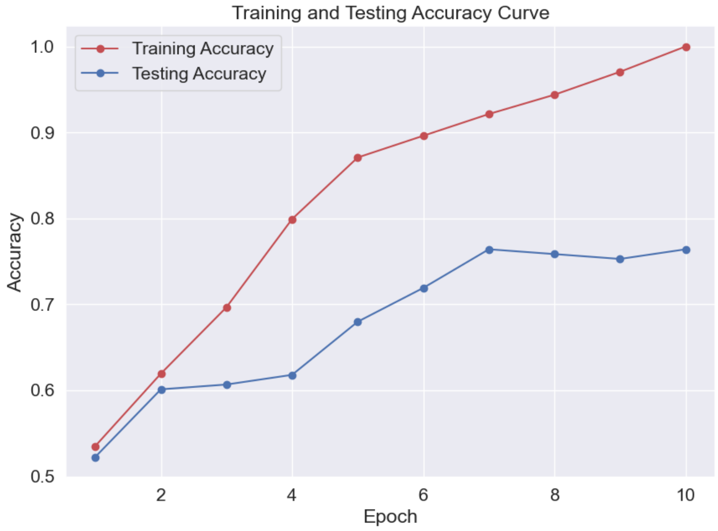

Fig. 12. Training and Testing Accuracy Curve of Random Forest

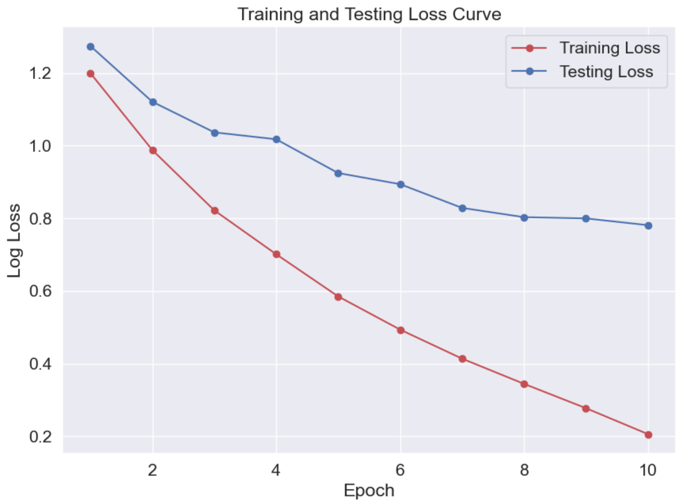

Fig. 13. Training and Testing Loss Curve of Random Forest

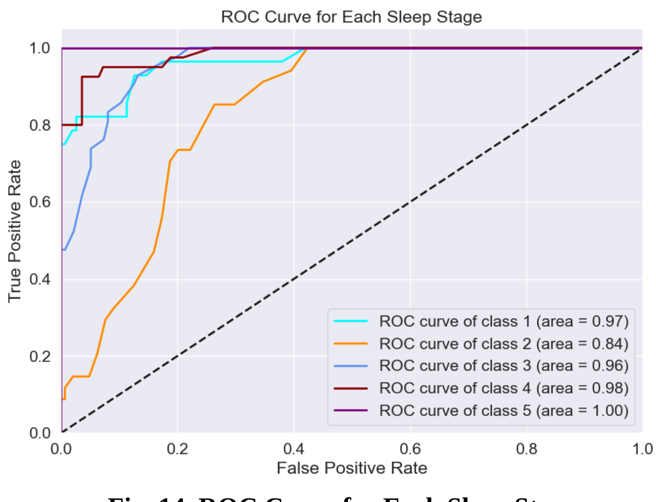

Fig. 14. ROC Curve for Each Sleep Stage

Fig.12 and Fig.13 illustrate the training and testing accuracy curve and loss curve. Similarly, Fig.14 presents the ROC curve obtained using the random forest classifier.

4.2 Gradient Boosting

Gradient boosting is an ensemble learning technique that constructs a series of decision trees in a sequential manner where each tree aims to correct the error of the previous tree by this process it iteratively aims to minimize the prediction error so gradient boosting makes a powerful choice over complex classification task such as sleep stage classification in this, we use 30 trees estimators, max depth 3, learning rate 0.1 and random state 42 for classification. The confusion matrix for gradient boosting is presented in Fig.15.Similarly the accuracy and loss curve for training and testing data is presented in Fig.16 and Fig.17 respectively.Fig.18 reports the ROC curve using gradient boosting techniques.

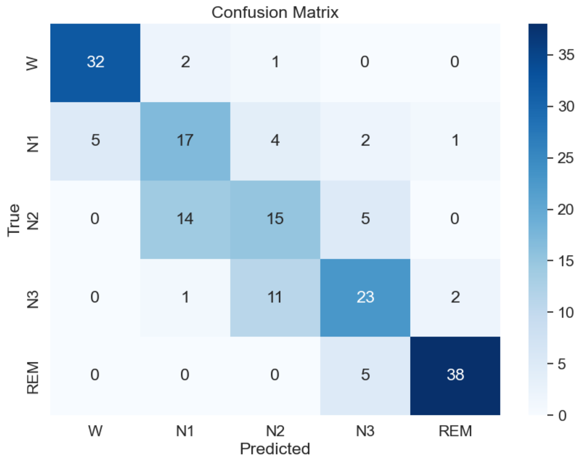

Fig. 15. Confusion Matrix Gradient Boosting

Fig. 16. Training and Testing Accuracy Curve of Gradient Boosting

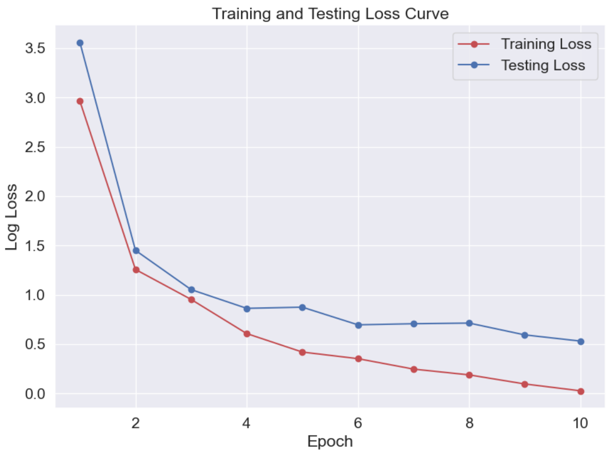

Fig. 17. Training and Testing Loss Curve of Gradient Boosting

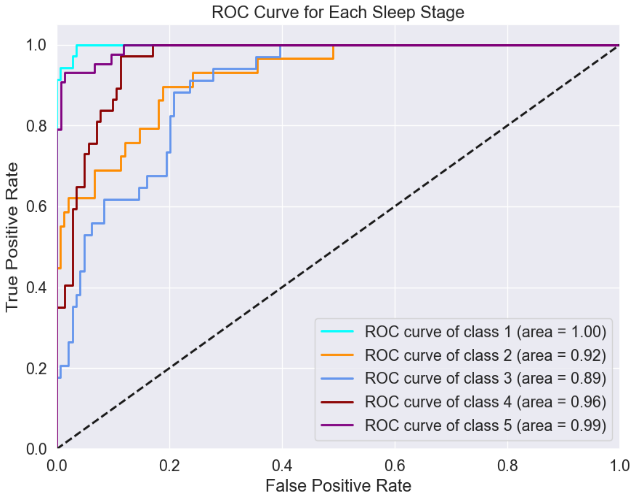

Fig. 18. ROC Curve for Each Sleep Stage

4.3 Ensemble Learning

In this ensemble learning, we have modified several changes first of all we handled the high dimensional EEG data with PCA we applied reduced feature space while retaining 95 % of the variance this is used to manage the overfit-ting particularly when working on complex data in an ensemble learning next we use same SMOT technique and we use combined multiple ensemble learning technique in we use combined of 3 techniques gradient boosting, random forest and support vector machine using soft voting classifier. soft voting averages the predicted probabilities from each classifier offering us more nuanced prediction over hard voting.

For classifier gradient boosting classifier, we use 50 estimator trees and a max depth of 3 at a learning rate of 0.1. in the second stack random forest classifier, we use 50 trees and a max depth of 3. In support vector classifier aims to find the optimal hyperplane to separate the classes here we use the radian basis function kernel to capture the nonlinear data patterns in the data. Fig.19 presents the confusion matrix using ensemble learning techniques. The accuracy and loss curve using training and testing data is presented in Fig.20 and Fig.21 respectively. Similarly, the ROC curve is presented in Fig.22.

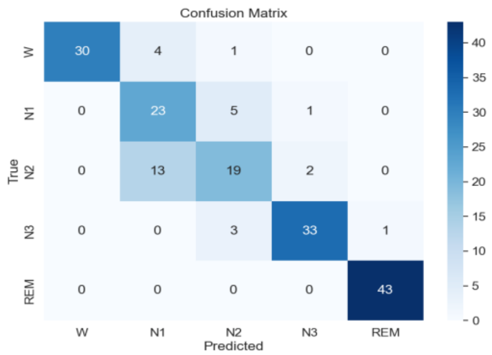

Fig. 19. Confusion Matrix Ensemble Learning

Fig. 20. Training and Testing Accuracy Curve of Ensemble Learning

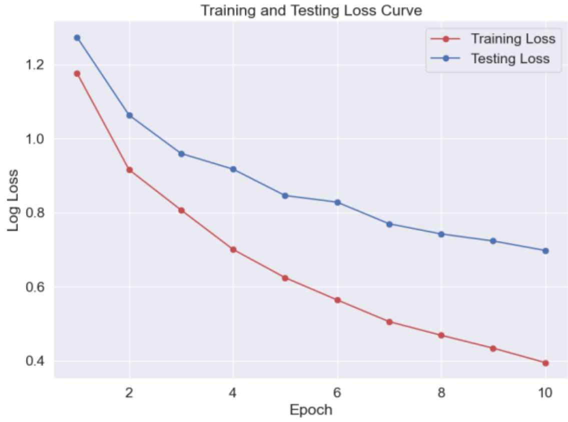

Fig. 21. Training and Testing Loss Curve of Ensemble Learning

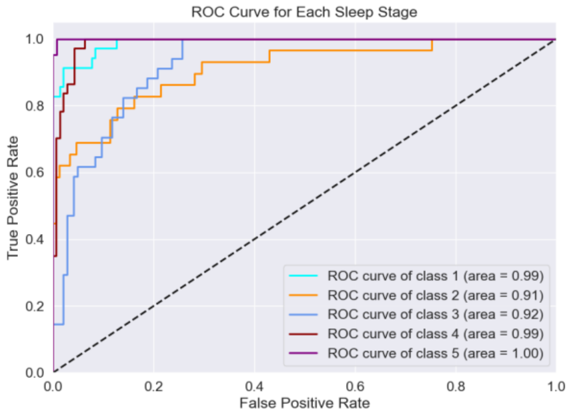

Fig. 22. ROC Curve for Each Sleep Stage

4.4 Bagging Classifier

We use a bagging classifier with a combination of PCA in which we reduce the feature space while retaining 95 % of the variance with the combination of SMOT Technique in the bagging classifier we use the max feature and max sample 1 and n estimator tree 30 for prediction now it has several advantages over the other techniques like reduce variance, robustness to overfitting, scalability and improved accuracy. The reason for improved accuracy is multiple week learners often result in better predictive performance for individual classifiers.Fig.23 presents the confusion matrix obtained from the bagging classifier.Fig.24 and Fig.25 present accuracy and loss curves using both the training and testing datasets.Fig.26 illustrates the ROC curve.

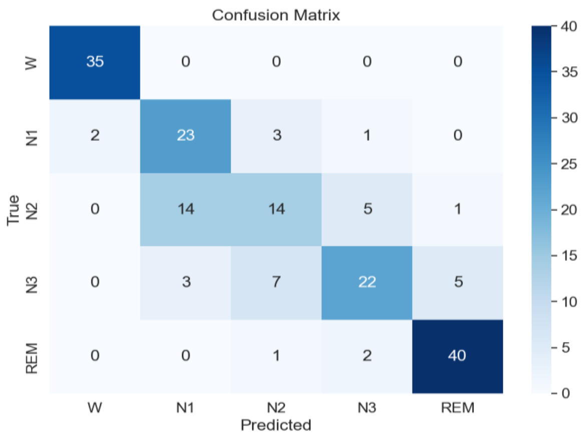

Fig. 23. Confusion Matrix Bagging Classifier

Fig. 24. Training and Testing Accuracy Curve of Bagging Classifier

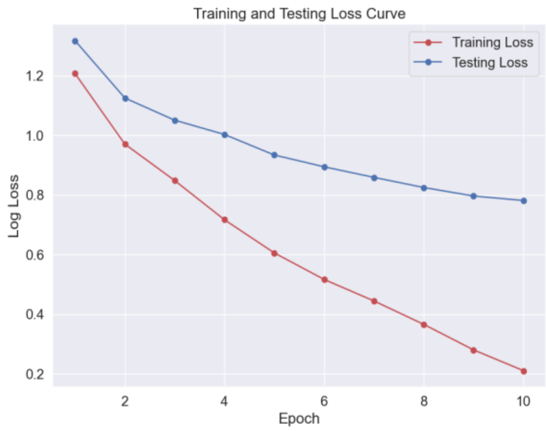

Fig. 25. Training and Testing Loss Curve of Bagging Classifier

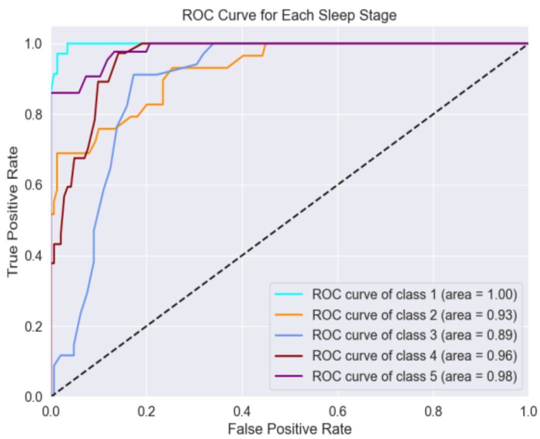

Fig. 26. ROC Curve for Each Sleep Stage

Table 1. Comparison of results of different obtained classifiers

Classifier	Accuracy	Precision	Sensitivity	F-Measure
Random Forest	0.764045	0.776646	0.769622	0.766187
Gradient Boosting	0.702247	0.686542	0.689402	0.686752
Ensemble Learning	0.831461	0.829502	0.820192	0.819052
Bagging Classifier	0.752809	0.736769	0.745939	0.733812

Ensemble Learning (Voting Classifier) After analyzing all the metrics, the best performance was given by the Ensemble Learning model, which had 83.15% accuracy and, with an F-measure of 81.91%, combined the predictions from several

estimators: Gradient Boosting, Random Forest, and Support Vector Classifier using soft voting, where the probability of prediction for each model was summed up to decide the final output. The main strengths of ensemble methods are leveraged from the diverse strengths of individual models, resulting in reduced variance and robustness to overfitting. Table 1 presents the performance metrics results using the different classifiers.
That is, high accuracy, precision, and sensitivity can be read as strong generalizations with a balanced prediction capability across different sleep stages. Weaknesses: While it achieved the best performance, it is computationally expensive to train multiple models and combine their predictions.Fig.27 presents the F1score performance.

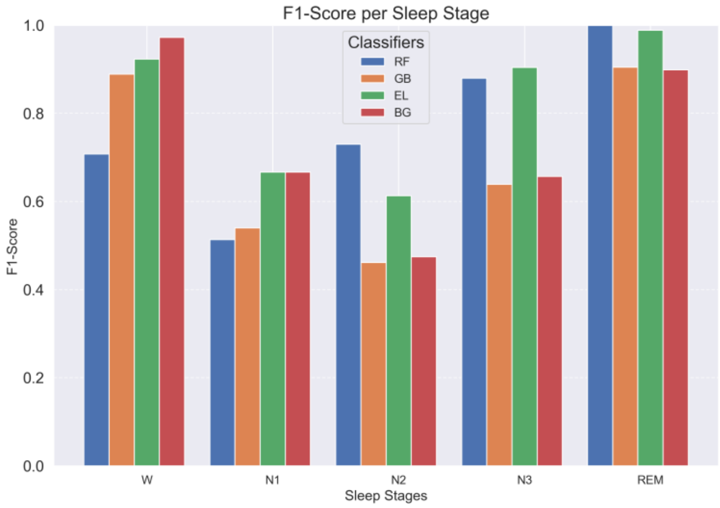

Fig. 27. Performance of F1 Score per class

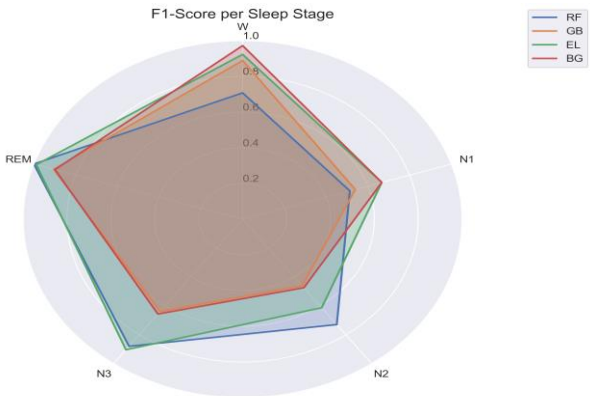

Fig. 28. Performance of F1Score Radar Score per class

Fig.28 presents the performance of the F1score with Radar chart using Random Forest After the Ensemble model, in terms of accuracy, it comes to the Random Forest classifier, 76.40%. While training, the Random Forest builds multiple trees, and the prediction is done by averaging the outputs of all the trees.

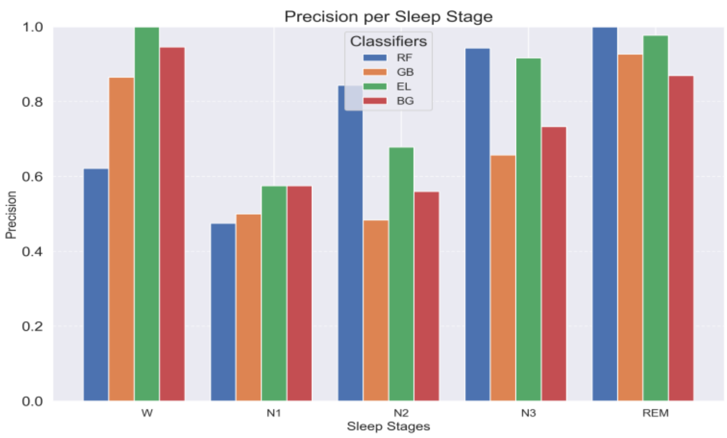

Fig. 29. Performance of Precision Score per class

Fig.29 illustrates the performance of the precision score per individual class.Fig.30 presents the radar graph representation of precision performances per stage. This will help to reduce the overfitting issues, which might come out, especially in noisy data like EEG signals. With an accuracy of 77.66% and an F-measure of 76.62%, Random Forest is a pretty good choice; though, it failed to outperform the Ensemble approach. Among all other options, the Ensemble approach has shown promise with much better performance measures of accuracy and F-measure. The strengths are as follows: High performance is related to accuracy and precision. Random Forest is highly interpretable and performs excellently in high-dimensional data. The weaknesses are Relatively lower sensitivity compared with Ensemble Learning; it might not catch minor patterns of certain sleep stages.

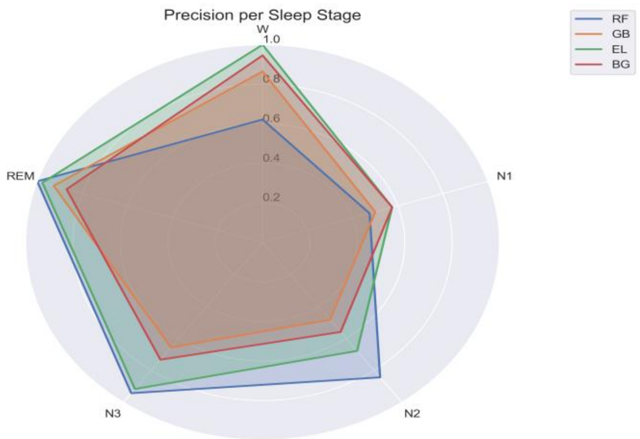

Fig. 30. Performance of Precision Radar Score per class

Bagging Classifier The best performance from the Bagging Classifier is 75.28%, which is competitive but a little behind that of Random Forest. A unicasting Bagging takes different random subsets of training data, trains multiple classifiers, and then aggregates them. Such a technique is quite effective in reducing the variance, but in this case, it has not performed any better than others. Precision at 73.68% and F-measure at 73.38% for the Bagging model demonstrate that while this model has been a good performer, it is more sensitive to variability in data compared with a Random Forest or Ensemble Learning model.Fig.31 and Fig.32 illustrate the recall performances per individual sleep stages.

Strengths: Generally good performance, with lowered variance. Bagging avoids overfitting, especially in models with inherently high variance. Weakness: While it is designed to reduce variance, Bagging underperformed compared to Random Forest and Ensemble Learning; which suggests that the bootstrapped datasets for Bagging may have limited its full potential in capturing the intricacies of the EEG data on this task.

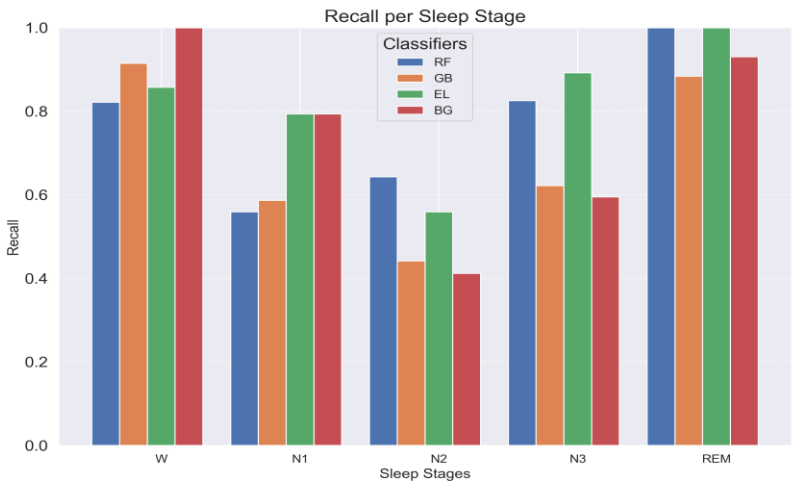

Fig. 31. Performance of Recall Score per class

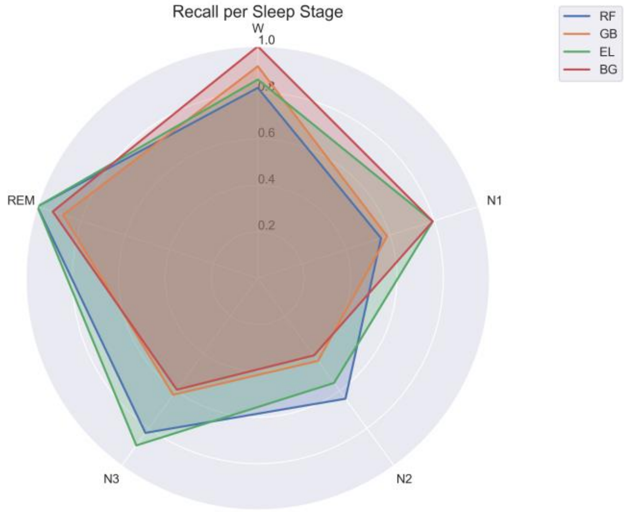

Fig. 32. Performance of Recall Radar Score per class

Gradient Boosting: The Gradient Boosting model was effective on many tasks but showed the lowest performance for this dataset, turning in an accuracy of 70.22% and an F-measure of 68.67%. The idea of Gradient Boosting is to train weak learners sequentially, with each subsequent learner trying to correct the mistakes of the previously trained one. However, this model was further handicapped in its current performance by the relatively lower sensitivity and precision measures among the investigated models.Fig.33 represents the impact of the classifier's performances of the obtained different classifiers in this proposed model.

Strengths: Gradient Boosting shines when the data presents somebody with complex patterns. More important is when the dataset requires an important refinement from a weak learner. It really can provide great predictive power with datasets that are very imbalanced or otherwise too hard to classify. Weaknesses: It had lower accuracy and sensitivity compared to other models. This could be because sequentially training weak learners on such complex high-dimensional EEG data is inherently hard. The general tendency of Gradient Boosting to overfit when its hyperparameters are not perfectly tuned might have contributed to its relatively poorer performance.

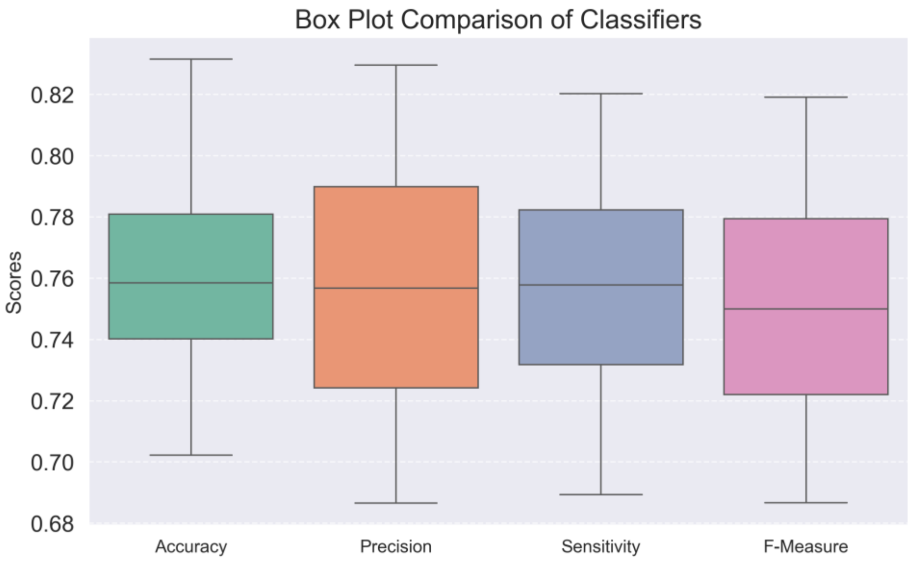

Fig. 33. Box Plot Comparison of Classifiers

Overview-Comparative Summary The performance was best by Ensemble Learning, which capitalized on the strengths of multiple classifiers to provide robust and accurate predictions across all sleep stages. However, Random Forest is only marginally worse than Ensemble Learning and still guarantees very high precision and F-measure; as a result, it is a good choice for sleep stage classification. The Bagging Classifier showed excellent performance, especially regarding the reduction of variance, though it did not outperform as much as Random Forest or Ensemble Learning. On the contrary, Gradient Boosting performed poorly, for it had demonstrated the lowest accuracy and F-measure, showing that the model struggles to handle the complexities of EEG sleep stage data. The comparisons of the proposed results with the existing work is presented in the Table 2.

Table 2. Sleep EEG Classification using Machine Learning

Studies	Classifiers	Dataset Used	Accuracy (%)
Zhang et al. (2019)	SVM	ISRUC	83.6
Zhang et al. (2018)	RF	Sleep-EDF	84.1
Koley & Dey (2018)	DT	DREAMS	82.1
Mousavi et al. (2017)	Naive Bayes	Sleep-EDF	81.0
Yildirim et al. (2018)	CNN	Sleep-EDF	84.5
Supratak et al. (2018)	RNN	MASS	83.9
Tsinalis et al. (2016)	CNN+RNN Hybrid	Sleep-EDF	83.0
Phan et al. (2017)	MLP	Sleep-EDF	82.7
Our Solution	Ensemble Learning	Sleep-EDF	85.0

In our experiment, we classify sleep stages using an ensemble learning approach that achieves a remarkable 85.0% accuracy on the Sleep-EDF dataset. Essentially, ensemble approaches are greatly beneficial as they take the strengths of multiple classifiers to the pool, thus making a model more robust as well as highly accurate. This is one method that counts by aggregating different classifier predictions and hence constitutes one powerful tool for complex tasks like sleep stage classification [1].

In this context, Zhang et al. (2019) used SVMs and obtained a figure of accuracy of 83.6% on the ISRUC dataset. Though SVMs are known for their superior performance in high-dimensional space, this performance is quite modest compared with our ensemble method. Additionally, Zhang et al. (2018) applied Random Forests on the Sleep-EDF dataset and achieved a value of 84.1%. Though random forests are robust, our ensemble method showed only small gains toward achieving the strength by combining multiple base learners to improve its predictive accuracy.

Further works show how various machine learning algorithms perform on sleep stage classification tasks. Koley and Dey (2018), for instance, report the application of a Decision Tree on the DREAMS dataset resulted in 82.1% accuracy probably due to the overfitting tendency of decision trees. Mousavi et al. (2017) reported 81.0% accuracy on a Naïve Bayes classifier on the Sleep-EDF dataset indicating simpler models cannot capture complex relationships between sleep data. The other methods include Yildirim et al. with the CNN approach at 84.5% and Supratak et al. with the RNN approach at 83.9%, which indicates that there is the capability to use deep learning methods, although they do not compete well with our ensemble method. Lastly, Tsinalis et al. (2016) proposed a CNN-RNN-based model which reached 83.0% accuracy while Phan et al. (2017) applied the multi-layer perceptron-MLP with an accuracy of 82.7%. To summarize, the ensemble approach has succeeded in yielding more recent performance in the following various studies and has proven that such an approach is efficient for sleep stage classification.

5 CONCLUSION

In this paper, we discussed classic machine learning techniques in use for sleep-stage classification using EEG data from the Sleep Physionet dataset. We made use of a Random Forest Classifier and SMOTE to deal with the issue related to class imbalance and achieved reasonable accuracy across all sleep stages involved. While Random Forest does provide a good baseline level of performance, mainly when applied to balanced datasets, it cannot keep up with the complexity and temporality inherent in EEG signals. Generally, it is challenging for this algorithm to decide, especially when the different stages such as N1 and REM are concerned. While these classical techniques such as Random Forest can capture some specific patterns within sleep stage data, more complex models, like deep learning methods, may be called for to manage the intricate and nonlinear characteristics of the EEG signals. This is more critical for deep learning models like CNNs and RNNs because of the ability to automatically learn features and capture temporal dependencies much more effectively. Hence, future work should be targeted for further augmentation in the performance on such complex datasets with enhanced models coupled with multimodal physiological data.

REFERENCES

[1] Alsolai, H., Qureshi, S., Iqbal, S. M. Z., Vanichayobon, S., Henesey, L. E., Lind-ley, C., & Karrila, S. (2022). A Systematic Review of Literature on Automated Sleep Scoring. In IEEE Access (Vol. 10, pp. 79419–79443). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2022.3194145

[2] Faust, O., Razaghi, H., Barika, R., Ciaccio, E. J., & Acharya, U. R. (2019). A review of automated sleep stage scoring based on physiological signals for the new millennia. In Computer Methods and Programs in Biomedicine (Vol. 176, pp. 81–91). Elsevier Ireland Ltd. https://doi.org/10.1016/j.cmpb.2019.04.032

[3] Tsinalis, O., Matthews, P. M., & Guo, Y. (2016). Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders. Annals of Biomedical Engineering, 44(5), 1587–1597. https://doi.org/10.1007/s10439-015-1444-y

[4] Ronzhina, M., Janoušek, O., Kolářová, J., Nováková, M., Honzík, P., & Prov-azník, I. (2012). Sleep scoring using artificial neural networks. In Sleep Medicine Reviews (Vol. 16, Issue 3, pp. 251–263). https://doi.org/10.1016/j.smrv.2011.06.003

[5] Biswal, S., Sun, H., Goparaju, B., Brandon Westover, M., Sun, J., & Bianchi, M. T. (2018). Expert-level sleep scoring with deep neural networks. Journal of the American Medical Informatics Association, 25(12), 1643–1650. https://doi.org/10.1093/jamia/ocy131

[6] Dutt, M., Redhu, S., Goodwin, M., & Omlin, C. W. (2023). SleepXAI: An ex-plainable deep learning approach for multi-class sleep stage identification. Ap-plied Intelligence, 53(13), 16830–16843. https://doi.org/10.1007/s10489-022-04357-8

[7] Han, H., Seong, M. J., Hyeon, J., Joo, E., & Oh, J. (2024). Classification and automatic scoring of arousal intensity during sleep stages using machine learning. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-023-50653-9

[8] Gaiduk, M., Serrano Alarcón, Á., Seepold, R., & Martínez Madrid, N. (2023). Current status and prospects of automatic sleep stages scoring: Review. In Bio-medical Engineering Letters (Vol. 13, Issue 3, pp. 247–272). Springer Verlag. https://doi.org/10.1007/s13534-023-00299-3

[9] Mousavi, S., Afghah, F., & Rajendra Acharya, U. (2019). Sleepeegnet: Auto-mated sleep stage scoring with sequence to sequence deep learning approach. PLoS ONE, 14(5). https://doi.org/10.1371/JOURNAL.PONE.0216456

[10] Zhao, C., Wu, W., Zhang, H., Zhang, R., Zheng, X., & Kong, X. (2024). Sleep Stage Classification Via Multi-View Based Self-Supervised Contrastive Learning of EEG. IEEE Journal of Biomedical and Health Informatics, 1–9. https://doi.org/10.1109/JBHI.2024.3432633

[11] Chen, Y., Lv, Y., Sun, X., Poluektov, M., Zhang, Y., & Penzel, T. (2024). ESSN: An Efficient Sleep Sequence Network for Automatic Sleep Staging. IEEE Journal of Biomedical and Health Informatics. https://doi.org/10.1109/JBHI.2024.3443340

[12] Jiang, C., Xie, W., Zheng, J., Yan, B., Luo, J., & Zhang, J. (2024). MLS-Net: An Automatic Sleep Stage Classifier Utilizing Multimodal Physiological Signals in Mice. Biosensors, 14(8). https://doi.org/10.3390/bios14080406

[13] He, M., Tang, M., Meng, L., & Liang, Z. (2024). TBSTSleepNet: Three-branch spectro-temporal bidirectional LSTM based attention model for EEG sleep stag-ing. Biomedical Signal Processing and Control, 97. https://doi.org/10.1016/j.bspc.2024.106695

[14] Zhang, W., Zhang, S., Wang, Y., Li, C., Peng, H., & Chen, X. (2024). A CNN-Transformer-ConvLSTM-CRF Hybrid Network for Sleep Stage Classification. IEEE Sensors Journal. https://doi.org/10.1109/JSEN.2024.3434404

[15] Ito, A., & Tanaka, T. (2024). SleepSatelightFTC: A Lightweight and Interpret-able Deep Learning Model for Single-Channel EEG-Based Sleep Stage Classifica-tion. https://doi.org/10.1101/2024.08.02.606301

[16] el Hadiri, A., Bahatti, L., el Magri, A., & Lajouad, R. (2024). Sleep stages de-tection based on analysis and optimisation of non-linear brain signal parameters. Results in Engineering, 23. https://doi.org/10.1016/j.rineng.2024.102664

[17] Ma, J., Lin, Q., Jia, Z., & Feng, M. (2024). ST-USleepNet: A Spatial-Temporal Coupling Prominence Network for Multi-Channel Sleep Staging. http://arxiv.org/abs/2408.11884

[18] Tsinalis, O., Matthews, P. M., & Guo, Y. (2016). Automatic Sleep Stage Scor-ing Using Time-Frequency Analysis and Stacked Sparse Autoencoders. Annals of Biomedical Engineering, 44(5), 1587–1597. https://doi.org/10.1007/s10439-015-1444-y

[19] Fiorillo, L., Puiatti, A., Papandrea, M., Ratti, P. L., Favaro, P., Roth, C., Bargio-tas, P., Bassetti, C. L., & Faraci, F. D. (2019). Automated sleep scoring: A review of the latest approaches. In Sleep Medicine Reviews (Vol. 48). W.B. Saun-ders Ltd. https://doi.org/10.1016/j.smrv.2019.07.007

[20] Chriskos, P., Frantzidis, C. A., Nday, C. M., Gkivogkli, P. T., Bamidis, P. D., & Kourtidou-Papadeli, C. (2021). A review on current trends in automatic sleep staging through bio-signal recordings and future challenges. In Sleep Medicine Reviews (Vol. 55). W.B. Saunders Ltd. https://doi.org/10.1016/j.smrv.2020.101377

[21] Satapathy, S.K., Loganathan, D. Automated classification of multi-class sleep stages classification using polysomnog-raphy signals: a nine-layer 1D-convolution neural network approach. Multimed Tools Appl (2022). https://doi.org/10.1007/s10420-022-13195-2

[22] Satapathy, S.K., Loganathan, D. Multimodal multiclass machine learning model for automated sleep staging based on time series data. SN COMPUT. SCI. 3, 276 (2022). https://doi.org/10.1007/s42979-022-01156-3

[23] Satapathy, S.K., Loganathan, D. Prognosis of automated sleep staging based on two-layer ensemble learning stacking model using single-channel EEG signal. Soft Comput 25, 15445–15462 (2021). https://doi.org/10.1007/s00500-021-06218-x

[24] Satapathy, S.K., Bhoi, A.K., Loganathan, D., Khandelwal, B., Barsocchi, P. Machine learning with ensemble stacking model for automated sleep staging using dual-channel EEG signal.Biomed Signal Process Control 69, 102898 (2021). https://doi.org/10.1016 /j.bspc.2021.10289

[25] Satapathy, S.K., Loganathan, D., Kondaveeti, H.K., et al. Performance analysis of machine learning algorithms on au-tomated sleep staging feature sets. CAAI Trans. Intell. Technol. 6(2), 155–174 (2021). https://doi.org/10.1049/cit2.12042

[26] Satapathy, S.K., Loganathan, D. Machine learning approaches with hetero-ogeneous ensemble learning stacking model for automated sleep staging. Int. J. Comput. Digit. Syst. (2022). https://doi.org/10.12785/ijcds/100109

[27] Satapathy, S.K., Loganathan, D. A study of human sleep stage classifica-tion based on dual channels of EEG signal using machine learning techniques. SN COMPUT. SCI. 2, 157 (2021). https://doi.org/10.1007/s42979-021-00528-5

[28] Satapathy, S.K., Loganathan, D. Automated classification of sleep stages using single-channel EEG: A machine learn-ing-based method. Int. J. Inf. Retr. Res. 12(2), 1–19 (2022). https://doi.org/10.4018/IJIRR.299941