Sleep Disorder Diagnosis Through Complex-Morlet-Wavelet Representation Using Bi-GRU and Self-Attention

Mubarak Albathan*
College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU),
Riyadh 11432, Saudi Arabia

***Abstract*—Sleep disorders pose notable health risks, impacting memory, cognitive performance, and overall well-being. Traditional polysomnography (PSG) used for sleep disorder diagnosis are complex and inconvenient due to complex multi-class representation of signals. This study introduces an automated sleep-disorder-detection method using electrooculography (EOG) and electroencephalography (EEG) signals to address the gaps in automated, real-time, and noninvasive sleep-disorder diagnosis. Traditional methods rely on complex PSG analysis, whereas the proposed method simplifies the involved process, reducing reliance on cumbersome equipment and specialized settings. The preprocessed EEG and EOG signals are transformed into a two-dimensional time-frequency image using a complex-Morlet-wavelet (CMW) transform. This transform assists in capturing both the frequency and time characteristics of the signals. Afterwards, the features are extracted using a bidirectional gated recurrent unit (Bi-GRU) with a self-attention layer and an ensemble-bagged tree classifier (EBTC) to correctly classify sleep disorders and very efficiently identify them. The overall system combines EOG and EEG signal features to accurately classify people with insomnia, narcolepsy, nocturnal frontal lobe epilepsy (NFLE), periodic leg movement (PLM), rapid-eye-movement (RBD), sleep behavior disorder (SDB), and healthy, with success rates of 99.7%, 97.6%, 95.4%, 94.5%, 96.5%, 98.3%, and 94.1%, respectively. Using the 10-fold cross-validation technique, the proposed method yields 96.59% accuracy and AUC of 0.966 with regard to classification of sleep disorders into multistage classes. The proposed system assists medical experts for automated sleep-disorder diagnosis.**

Keywords*—*Deep learning; complex morlet wavelet; bidirectional gated recurrent unit; sleep stage detection; multistage sleep disorder; ensemble-bagged tree classifier

I. Introduction

Sleep is an essential concern for human health. It serves as a fundamental factor in physical and mental wellness. It is vital in memory consolidation, cognitive functions, cellular regeneration, and metabolic-brain-waste elimination [1]. Therefore, abnormalities in normal sleep patterns can cause many disorders, such as insomnia, narcolepsy, and sleep apnea disorder. Each sleep condition affects each individual's health differently, often inducing daytime weariness, cognitive impairment, cardiovascular disease, and mental health difficulties [2]. Therefore, accurate sleep-disruption assessment and therapy are crucial for overall health and quality of life. Polysomnography (PSG) is the most reliable sleep problem diagnosis method, but it involves overnight stays at medical institutions, which can be resource-intensive and uncomfortable for patients. The need for more accessible and user-friendly sleep problem diagnosis and analysis is evident, considering these restrictions and the discomfort it brings to patients.

This paper presents a unique automated sleep-disorder-detection approach employing EOG and EEG signals [4]. Sleep disruption monitoring using EOG and EEG is simple and less invasive. EOG records eye movements, which distinguish sleep phases, whereas EEG records muscle activity [5]. Instead of sophisticated PSG analysis, the suggested technique streamlines the operation, minimizing the need for expensive equipment and particular settings. A unique sleep disorder classification method uses sophisticated signal processing and machine learning. Complex signal processing approaches like Morlet-wavelet transformations and machine learning models with bidirectional gated recurrent units and self-attention layers make sleep diagnosis easier and more accurate, instilling a sense of confidence. This approach is appropriate for home monitoring since it reduces equipment and simplifies diagnosis. EEG and EOG signals detect sleep problems well. By identifying and assessing the most important signal information, diagnostic accuracy techniques typically outperform traditional PSG. Finally, EOG and EEG signal analysis improves sleep issue detection and efficiency in this study, providing reassurance of its accuracy and efficiency. It improves efficiency, accessibility, and patient comfort while maintaining the diagnostic integrity of traditional PSG.

Standard machine learning (ML)–based sleep-disorder detection involves the use of concepts such as support vector machines (SVMs), decision trees, and k-nearest neighbor (k-NN) algorithms. The study of [6] demonstrated the efficacy of SVMs in accurately classifying sleep apnea (accuracy rate = 85%) using EMG data. They highlighted the capability of machine learning in sleep disorder identification; however, handling data with a high number of dimensions was difficult. the research of [7] employed RF algorithms to differentiate various sleep disorders, such as insomnia and narcolepsy, by analyzing EOG signals. In this regard, they achieved an impressive accuracy rate of 89%, showcasing the efficacy of ensemble techniques. Despite exhibiting positive results, traditional machine learning–based approaches have several constraints. Feature selection is a primary issue in this regard, which often involves human participation and is, therefore, vulnerable to bias. To address this challenge, Aboalayon et al.

*Corresponding Author.

[8] effectively employed an automated feature selection technique, which substantially enhanced the involved model's performance (by 5%). However, such techniques faced challenges in terms of their comprehensibility and ability to be applied to new, unexpected data.

The advent of deep learning has considerably transformed sleep disorder detection [9]. Convolutional neural networks (CNNs) and RNNs have gained popularity because they can automatically extract features and detect patterns in time-series data. A CNN was employed herein to examine EMG data for rapid-eye-movement (REM) sleep behavior disorder (RBD) identification, yielding a precision of 95%. This accomplishment indicates a substantial improvement compared with traditional sleep-disorder-identification models. Further, LSTM was used to determine the time-dependent patterns of EOG signals to identify sleep stages. Notably, LSTM yielded a remarkable classification rate accuracy rate of 92% in this regard in a previous study. Despite their impressive accuracies, deep learning algorithms require large amounts of data and considerable computational resources, among other challenges. Moreover, these algorithms exhibit an issue about interpretability owing to the lack of transparency. Therefore, studies (e.g., an experimental study by Meridian et al. [10]) have been exploring hybrid models that involve a combination of CNNs with LSTM networks. This combination enables the usage of both spatial and temporal properties to improve model transparency and efficiency.

Conventional sleep problem diagnosis methods sometimes need resource-intensive polysomnography (PSG) and uncomfortable conditions. The study addresses these issues. While promising, current deep learning (DL) and machine learning (ML) methods are sometimes computationally intensive, require extensive hyperparameter tuning, and are often restricted to particular sleep phases or classes, resulting in solutions that are not very generalizable. Furthermore, the high level of data complexity and the widespread usage of hyperparameters in many existing algorithms cause overfitting, further restricting their application to various datasets.

A unique automated method that combines EEG and EOG data to diagnose sleep difficulties is presented in the paper. Combining advanced signal processing with a complex-morlet-wavelet (CMW) transform and a Bi-GRU with a self-attention layer simplifies the diagnostic process. This strategy is more accessible for home monitoring since it involves less sophisticated installations and heavy-duty equipment. The recommended technique also reduces overfitting and processing requirements while boosting sleep disorder classification accuracy, making it a more practical and complete solution for real-world applications.

This research substantially contributes to the field of sleep disorder through the following:

1) An innovative approach is introduced, involving complex-morlet-wavelet (CMW) transform–based feature extraction from preprocessed EOG and electroencephalography (EEG) signals. This method significantly enhances the accuracy of multistage sleep-disorder identification.

2) A bidirectional gated recurrent unit (BiGRU) with a self-attention layer is employed for feature extraction, followed using an ensemble-bagged tree classifier (EBTC) for precise multistage-sleep-disorder classification. This methodology ensures accurate delineation of complex sleep patterns by leveraging temporal insights and robust ensemble methods.

3) The effectiveness of the proposed method is robustly demonstrated in a real-world setting, offering a practical and reliable solution for home-based, patient-friendly sleep-disorder monitoring.

The remainder of this paper is organized as follows: Section II presents the relevant past studies. Section III comprehensively explains the approaches used herein, encompassing a CMW transform, feature extraction, and sleep disorder classification. Section IV presents the outcomes achieved regarding the categorization of people into healthy individuals and individuals with sleep disorders and the classification of sleep disorders into seven categories. Section V concludes the work.

II. Literature Review

A previous study classified sleep disorders using a novel machine-learning model that combined EEG, chin EMG, and dual-channel EOG [11]. The authors used the best orthogonal filterbank and Tsallis entropies to obtain high classification accuracies when considering the Sleep Heart Health Study (SHHS) database (90.7% for SHHS-1 and 91.8% for SHHS-2), achieving excellent automated sleep-disorder classification. Jarchi et al. [12] aimed to diagnose breathing- and eye-movement-related sleep disorders using electrocardiography (ECG) and EMG by developing a deep learning framework that yielded a mean accuracy of 72% in classifying people into four groups—healthy individuals and individuals with various sleep disorders (obstructive sleep apnea (OSA), restless leg syndrome (RLS), or both). This demonstrated the capacity of ECG and EMG in diagnosing sleep disorders. Meanwhile, Sharma et al. [13] introduced an automated technique for sleep disorder identification that involves analyzing EOG and EMG signals. A biorthogonal filter bank with Hjorth parameters was employed, which yielded a high overall accuracy of 94.3%. This technique was recognized for its effectiveness in at-home monitoring of different sleep disorders. Sekkal et al. [14] compared eight classic machine-learning techniques with a feed-forward neural network for sleep disorder identification and discussed the pros and cons of various sleep stage classifiers.

Sharma et al. [15] exploited EEG data to diagnose sleep problems. Using Hjorth parameters and an ensemble-boosted tree classifier, they classified sleep disorders with 99.2% accuracy. This method helps clinicians detect sleep problems. Rahman et al. [16] examined automated sleep-stage evaluation using EOG data. They used discrete-wavelet-transform EOG data to improve S1-sleep-stage detection over previous EOG-based approaches. Pei et al. [17] created a successful deep-learning model for sleep phases utilizing biological cues. Combining CNNs with gated recurrent units (GRUs) yielded a more versatile model than previous cutting-edge models.

An automated sleep-stage approach using EEG, EOG, and EMG was developed by Satapathy et al. [18]. The system found linear and nonlinear characteristics with good classification accuracy and diagnosed sleep disorders. A novel sleep staging approach used EOG instead of EEG for practicality. This technique has 81.2% and 76.3% sleep-staging accuracy utilizing a two-scale CNN and RNN. Chambon et al. [20] used PSG data to characterize sleep phases using deep learning without explicitly designing characteristics. A fair use of channels and temporal data gave the model excellent classification performance. EEG-based sleep stage categorization using PSG analysis was advanced by several research [21–25]. This research showed that several machine learning methods and physiological signal combinations produced excellent accuracy and showed the benefits of multimodal signal processing.

In another research [26], a restricted PSG montage classified the sleep phases of 106 people—53 with RBD and 53 healthy. This was done using an RF classifier with 156 EEG, EOG, and EMG characteristics. RBD was detected using muscle atonia measurement and sleep architecture characteristics. The model attained Cohen's Kappa score of 0.62 for sleep staging and 96% RBD detection accuracy, demonstrating the benefits of sleep architecture and transitions. Malafeev et al. [27] developed a three-dimensional (3D) CNN for sleep stage categorization using several channels and EEG, EMG, and EOG inputs. Time, frequency, and time-frequency characteristics were sent to the 3D CNN. Three-dimensional convolutional layers created intrinsic relationships between biosignals and frequency bands, while two-dimensional layers obtained frequency correlations. The model identified significant channels and frequency bands throughout sleep phases using partial-dot-product attention layers and an LSTM unit. This model also achieved classification accuracies of 0.832 and 0.820 on the ISRUC-S3 and S1 datasets. These findings showed the model detected sleep phases reliably and effectively.

Cooray et al. [28] proposed "quasi-normalization" for feature normalization using the ISRUC-Sleep dataset. An RF algorithm sorted the data into five sleep states. Using leave-one-out cross-validation, EOG and EMG data were integrated to achieve Cohen's kappa value of 0.749 and 80.8% accuracy. The results matched the American Academy of Sleep Medicine standards. Electrooculography and electromyography may be as effective as electroencephalography at identifying sleep phases. Another research [29] studied sleep phases in 123 suspected sleep disorder patients using a BiLSTM network. The model received multivariate time-series heart rate, breathing rate, and body movement frequency. With an accuracy of 71.2%, Cohen's κ coefficient of 0.425, and an F1 score of 0.650, the model effectively classifies sleep phases using minimal physiological cues.

Morokuma et al. [30] focused on EEG and EOG signals and developed a deep CNN architecture for automated sleep-stage classification. Its performance was evaluated against human expert agreement, with CNN considerably outperforming recent single-EEG-channel approaches. The study highlighted the crucial role of network depth in achieving high classification accuracy. Another study [31] targeted sleep-wake detection in OSA patients using single-channel ECG signals. The heart rate variability signals were derived, and features were classified using decision trees, SVMs, and ensemble classifiers. The model achieved accuracies of 81.35% with three features and 87.12% with ten features, suggesting its utility in the OSA diagnosis. An automated deep nine-layer one-dimensional CNN (9L-1D-CNN-SSC) for multiclass sleep staging was also developed [32]. The model was tested on ISRUC-Sleep subgroup datasets and achieved an accuracy of up to 99.50% in classifying sleep stages with different signal combinations, indicating its applicability for clinical use. Satapathy and Loganathan [33] developed a dual-modal, multiscale deep neural network for sleep staging that used EEG and ECG signals. When tested on the MIT-BIH PSG dataset, the model achieved high accuracy rates of 80.40%–98.84% in classifying different sleep stages. This shows that combining EEG and ECG signals for sleep analysis yields accurate results.

Zhao et al. [34] addressed the limitations of automated sleep-staging systems using portable EEG headbands by developing a deep-learning model using convolutional and long-term memory layers. The model achieved validation accuracies of 74% on headband data and 77% on PSG data, demonstrating its potential in ambulatory sleep assessments. SleepPrintNet [35] was also introduced to capture the SleepPrint in a physiological time series for sleep staging. It yielded higher accuracy on the MASS-SS3 dataset than baseline models because it used EEG, EOG, and EMG features along with temporal, spectral, and spatial features. This approach underscored the value of multimodal feature integration in sleep stage classification.

Several studies [26–36] collectively represented a wide array of methodologies, ranging from RF classifiers to BiLSTM networks and CNNs. These methods were tested on various datasets such as the SHHS and ISRUC-Sleep datasets, achieving considerable advancements in sleep stage classification, disorder diagnosis, and automated sleep-staging systems. Many of these studies highlighted the effectiveness of combining EEG, EOG, EMG, and ECG signals, emphasizing on the efficiency of feature extraction and selection in improving accuracy and robustness. The comparative analysis of these systems is shown in Table I.

TABLE I. COMPARISON OF METHODOLOGIES, RESULTS, AND LIMITATIONS PROPOSED IN STUDIES

Reference	Methodology	Results	Limitations
[11]	A machine learning model using EEG, chin EMG, and dual-channel EOG	90.7% accuracy on SHHS-1 and 91.8% on SHHS-2	three sleep classes in SHHS-1, and five classes in SHHS-2 datasets, none-generalized.
[12]	Diagnosing sleep disorders using ECG and EMG and a deep learning framework	Mean accuracy of 72% and weighted F1 score of 0.57	Four-sleep classes, Computational expensive and huge hyper-parameters sitting.
[13]	An automatic detection system for sleep disorders using EOG and EMG signals	Accuracy of 94.3%	Limited five-sleep classes, and three sleep stages.
[15]	Identification of sleep disorders using EEG and EBTC	Classification accuracies up to 99.2%	Limited four-sleep classes and used only one CAP dataset so non-generalize solution.

[17]	A deep learning method using CNNs and GRUs for sleep stage identification	Accuracy of 83.15% and kappa of 0.76	Limited five-sleep stages and not classes, and computationally expensive
[18]	An automated sleep-staging system using EEG, EOG, and EMG signals and an RF classifier	Accuracy of 98.99%, 98.75%, 98.17%, and 99.14% with respect to sleep stage	five-sleep states, non-sleep classes, and computationally expensive
[19]	A sleep staging approach using EOG and two-scale CNNs and RNNs	Accuracy of 81.2% of two-scale CNNs and 76.3% of RNNs.	Limited sleep classes, not generalized, and required huge hyper-parameters.
[20]	Convolutional deep learning approach and gradient boosting for sleep stage classification using PSG signals	Accuracy of approx. 80%.	Huge hyperparameters, Five-sleep stages
[21]	An efficient technique for sleep stage classification based on EEG signal analysis	RF algorithm achieved a high accuracy of 97.8%	Limited dataset, Overfitting due to RF and required stopping critria, three-stages of sleep disorder.
[22]	SleepEEGNet for automated sleep-stage annotation using single-channel EEG and BiRNN.	Accuracy of 84.26%, F1-score of 79.66% and $\kappa = 0.79$.	Limited sleep stage, tested on limited dataset
[23]	A deep learning model for sleep staging in children using EEG, EOG, and chin EMG	Cross-validated accuracy of 84.1%	Limited sleep classes and computationally expensive
[24]	A deep learning model for sleep staging using multiple PSG signals and 2D CNNs and LSTM modules	Sleep-EDF: Acc-0.86, K-0.81	Limited sleep classes, not generalized
[25]	Sleep staging using Relief, AdaBoost with RF	accuracy of 97.96%	Limited sleep classes, not generalized, and classifier overfitting
[26]	Sleep stages using EEG, EMG, and EOG signals and CNN-LSTM	Accuracy of 0.832 on ISRUC-S3 and 0.820 on ISRUC-S1	Classifier overfitting, and Computationally expensive
[27]	Sleep stage classification using 3D-CNN	Accuracy of 0.832, F1-score of 0.814 and kappa of 0.783 on ISRUC-S3	Required huge hyper-parameters sitting and computationally expensive due to huge epochs.
[28]	EOG and EMG data are utilized to predict multistage sleep by using RF classification	Accuracy of 92%	Limited sleep classes and classifier overfittig, limiated dataeset so no generalize solution
[29]	A ”quasi-normalization” method with RF classifier	Accuracy of 84.7%	Not generalized, overfitting
[30]	Polysomnography (PSG) data with BiLSTM classifier for detection of sleep stages	Accuracy of $71.2 \pm 5.8\%$, and F1 score of 0.650 ± 0.083	Only sleep stages and no multiclass solution.
[31]	Detection of human sleep EEG and EOG signals with CNN architecture.	F1-score of 77%	Limited sleep classes, not generalized
[32]	An ensemble technique based on three classifiers: DT, kNN and SVMs.	Sensitivity and specificity values of 0.90 and 0.85, respectively.	Limited two-sleep classes, not generalized
[33]	A 9L-1D-CNN-SSC model for sleep staging with different signal combinations	Classification accuracies up to 99.50% for various sleep stages	Limited sleep classes, not generalized, overfitting, and computationally expensive
[33]	A dual-modal multiscale deep neural network using EEG and ECG signals	Accuracies between 80.40% and 98.84% for different sleep stages	Overfitting and computationally expensive
[34]	A deep learning model with convolutional and LSTM layers for EEG headband data	74% accuracy on headband data and 77% on PSG data	Limited sleep classes, not generalized, overfitting, and computationally expensive
[35]	SleepPrintNet integrating EEG, EOG, and EMG signal features	Outperformed baseline models in accuracy on the MASS-SS3 dataset	Limited dataset utilized.

III. Methodology

The multilayer sleep-disorder classification system's systematic flow diagram is shown in Fig. 1. The novel method categorized sleep disorders using preprocessed EOG and EEG information. Normalization was done to standardize these signals for analysis. Next, a bandpass filter reduces noise and frequencies to increase signal quality, which is important for detecting sleep disorder symptoms. Next, the complex morlet wavelet (CMW) transform was utilized to extract features from EOG and EEG data for reliable disease categorization. A BiGRU with a self-attention layer extracted characteristics, and an EBTC automatically found sleep problems. Due to its efficiency in processing time-series data, the GRU, a kind of RNN, was utilized to evaluate EOG and EEG temporal patterns. By aggregating estimates from several decision tree models, the EBTC improved its accuracy and applicability. Finally, the BiGRU and EBTC models were integrated to categorize the data using voting, average probability, or a more complex meta-classifier. Using 10-fold cross-validation, the system's performance was rigorously assessed to ensure its efficacy and robustness in real-world circumstances. The suggested method improves sleep problem classification and shows the potential of signal processing and machine learning for medical diagnosis.

A. Data Acquisition and Augmentation

Define abbreviations and acronyms the first time they are The CAP Sleep database [37], provided by PhysioNet [38], containing PSG recordings from 108 people, was used as the primary data source. The EEG and EOG signals were collected and analyzed from this database.

The EEG and EOG signals are collected from individuals while they are asleep to detect sleep stages and classify sleep disorders. The EEG and EOG signals were comprehensively distributed (Table II) to evaluate the effectiveness of the proposed system in diagnosing different sleep disorders such as insomnia, narcolepsy, nocturnal frontal lobe epilepsy (NFLE), periodic leg movement (PLM), RBD, and sleep-disordered breathing (SDB) as well as healthy individuals. Fig. 2 shows a visual representation of signals from each group.

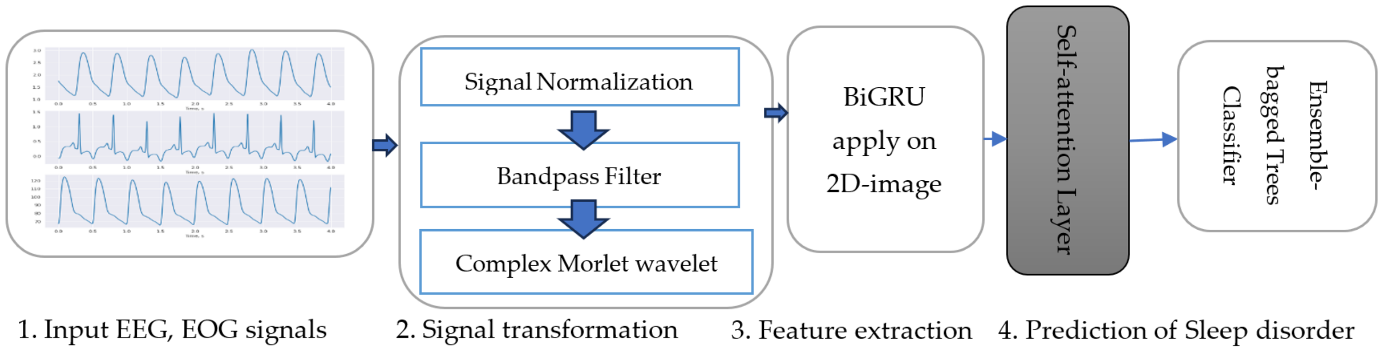


Fig. 1. Systematic flow diagram of the proposed multilayer sleep-disorder classification system.

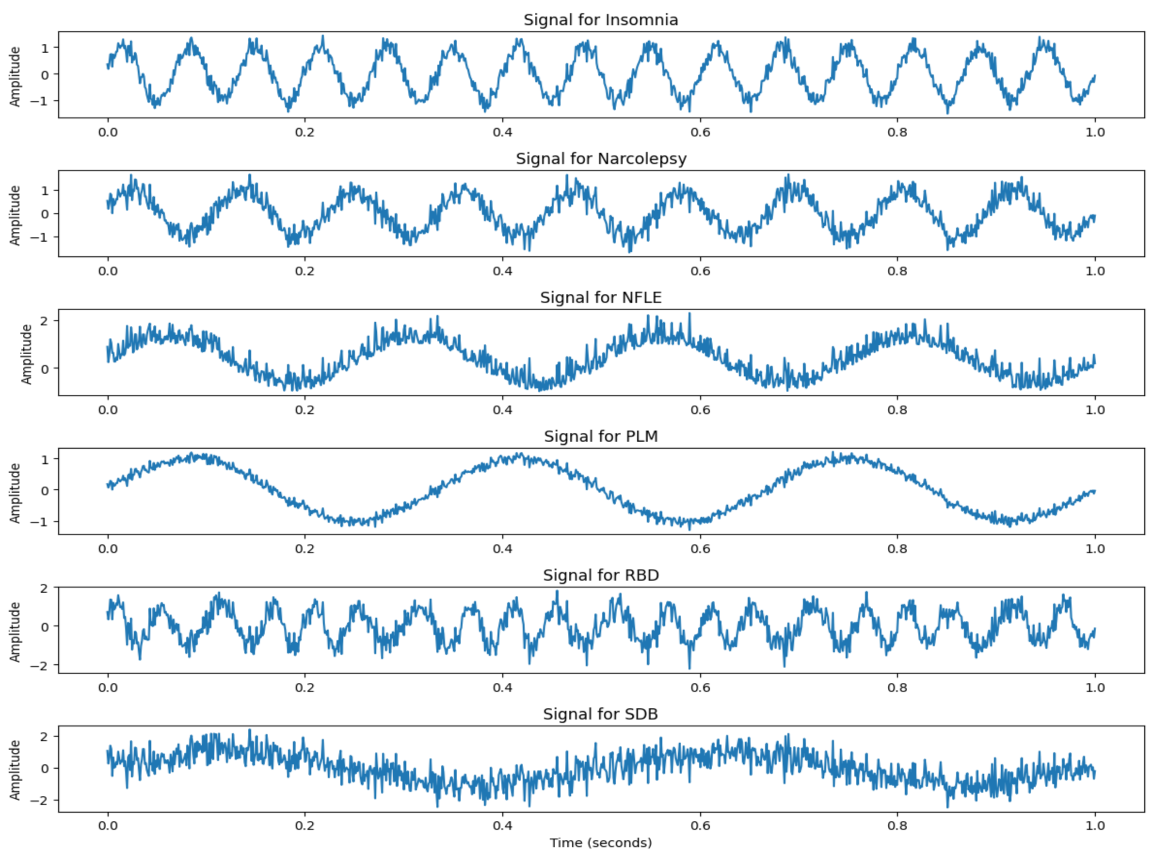


Fig. 2. Input EEG and EOG signals for identifying sleep disorders.

To standardize the sample, the dataset was subjected to data augmentation (Table II) using the synthetic minority over-sampling technique (SMOTE). SMOTE is used to rectify imbalanced datasets, particularly medical data in sleep studies, where certain classes are inadequately represented. It generates artificial, yet believable, examples using the available data from the underrepresented class. During the procedure of SMOTE, the dataset was assumed to contain EEG and EOG signal features, referred to as "features, "along with their corresponding labels, referred to as "labels." These labels classify each group of features into distinct sleep disorders or stages. The dataset was initially divided into 70% training and 30% testing datasets. SMOTE was used only on the training set to prevent synthetic data from affecting the model evaluation. It generated additional samples for classes that were not well represented, thus equalizing the distribution of classes, as balanced datasets improve model performance, particularly in classification tasks.

The training dataset (X-res and Y-res) contained the original and newly synthesized samples. This dataset was then used to train a machine-learning model. Notably, the model acquired knowledge from a dataset that provided a more equitable representation of all categories, thus mitigating its inclination toward the dominant category. The model's effectiveness was assessed using the initial, unmodified testing dataset. This approach ensures a complete evaluation of SMOTE's effectiveness as it accurately determines the effect of SMOTE on the model's ability to separate different sleep states.

TABLE II. EEG AND EOG SIGNAL DISTRIBUTION USED TO TEST THE SYSTEM PERFORMANCE

Sleep Stage	EEG	EOG	Total EEG and EOG	Data Augmentation
Insomnia	3800	1200	5000	2500
Narcolepsy	1300	1400	2700	2500
NFLE	3200	2000	3400	2500
PLM	1300	1000	2300	2500
RBD	5000	2000	7000	2500
SDB	200	100	300	2500
Healthy	400	200	600	2500

B. *Signal Transformation*

This multiclass sleep disorder prediction research uses the complex morlet wavelet transform (CMWT) [39] because it provides amplitude and phase information, unlike spectrograms, which indicate magnitude, and scalograms, which reveal phase information. While scalograms are superior for non-stationary signals, they cannot match the CMW transform. Spectrograms provide a broader perspective of power distribution. Arranging these normalized features creates the 2D stack picture, which is then fed into machine learning systems like the bidirectional gated recurrent unit (BiGRU), which maintains signal characteristics' spatial and temporal correlations.

PyWavelets were used to analyze a continuous wavelet transform (CWT). CWT is a robust time-frequency analysis method that can analyze signals at multiple scales or frequencies. This study employed the complex morlet wavelet (CMW), which analyzes nonstationary biological data via temporal and frequency localization. The CWT captured both high-frequency and low-frequency components of each simulated disorder signal by decomposing it into distinct scales. Each row and column of the coefficient matrix represented a frequency range and time point, respectively. The coefficients were subsequently represented using a heatmap, which displayed the frequency characteristics of the signal over time. The color intensity of the heatmap corresponded to the magnitude of the signal across different frequencies, providing valuable information about the distinctive patterns associated with each sleep problem.

Thus, the CMW is ideal for feature extraction in biological signal processing. By explaining the complicated time frequency features of EEG and EOG signals, this study presents in-depth grouping sleep disorders into different categories of identifying unique brain patterns. The proposed approach conforms with current research practices in biomedical engineering and computational neuroscience, focusing on advanced signal processing techniques to understand complicated physiological events. Fig. 3 displays a visual representation of each sleep-disorder type.

C. *Feature Extraction Using Bigru-Attention*

A sophisticated neural network structure, namely a BiGRU [40] with a self-attention mechanism, was used to mine the time-series data, such as EEG and EOG signals. This study presents a promising method that takes the concatenated wavelet features of the EEG and EOG signals to create a unified 2D representation of these features. Owing to its bidirectional nature, BiGRU analyzed patterns in forward and reverse directions throughout time and revealed the intrinsic temporal dynamics in signal data. Fig. 4 shows the BiGRU architecture, wherein signal data are used to find and extract relevant features indicating the essential traits of different sleep stages or disorders.

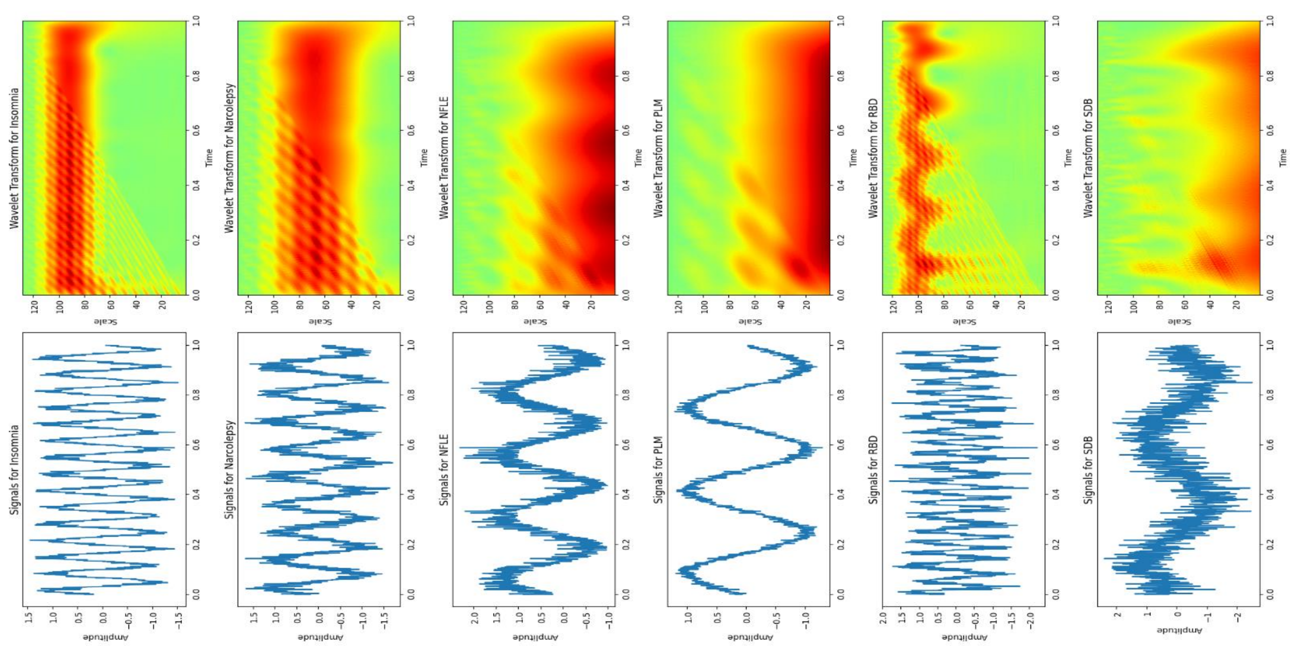


Fig. 3. Complex-morlet-wavelet transform using electroencephalogram signals of sleep disorders such as insomnia, narcolepsy, NFLE, PLM, RBD, and SDB.

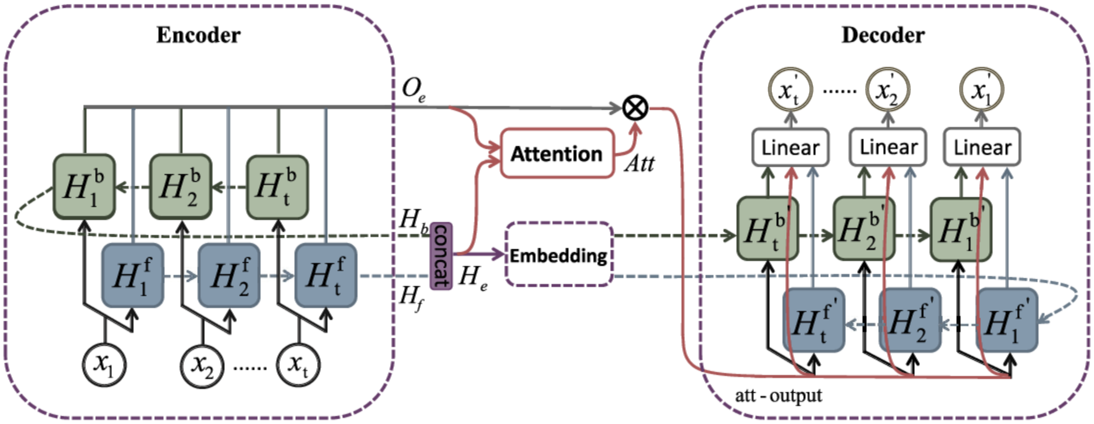


Fig. 4. Architecture of BiGRU with a self-attention layer.

EEG and EOG signals can be better analyzed by combining a self-attention mechanism with a BiGRU model, particularly when the signals change into a complex wavelet domain. The self-attention mechanism is a highly efficient neural network component, allows the model to prioritize different areas of the input data, making it more contextually aware. The sophisticated morlet wave transform further refines this by dividing EEG and EOG data into high- and low-frequency patterns, providing a deeper understanding of sleep stages and difficulties. The BiGRU model, in turn, uses these modified signals with rich time-frequency information for forward and backward analysis, capturing temporal correlations and patterns. This self-attention mechanism, with its efficiency, helps the model focus on essential signal alterations. It prioritizes segments that provide task-relevant information, such as sleep issues or phases, with high performance and accuracy.

Complex wavelet transformations, BiGRU, and self-attention mechanisms create a robust signal-processing paradigm. Precision frequency information was added to signals using the wavelet transform. The BiGRU neural network caught temporal patterns, and the self-attention mechanism focused on the most critical parts. When utilized together, they can extract crucial and relevant information from EEG and EOG data, making sleep studies and associated research more reliable and valuable. This strategy increases sleep study categorization and predictions by deepening physiological signal understanding.

EEG and EOG signals are converted to CMW for sleep disorder research and fed into the BiGRU model. Time-frequency analysis often uses the morlet wavelet because it splits initial signals into signals with various frequencies to optimize temporal and spectral localization. Thus, the BiGRU model learns from scale patterns. This model may detect small brain activity and eye movement changes that signal sleep phases and issues. This system employs a CMW transform and BiGRU model to use the BiGRU model's comprehensive time-frequency signal representation and powerful sequence modeling. By highlighting essential frequency components, the wavelet transform improves signals. The BiGRU then extracts key characteristics from this modified data, providing a robust collection of features for analysis or classification. This method handles EEG and EOG signal complexity and volatility well, making it suitable for advanced sleep investigations and diagnostics.

A BiGRU is a better version of the regular GRU developed for the model to obtain information from states preceding and succeeding the unit in a sequence. This is particularly advantageous in situations where the overall context of the entire sequence is crucial for making accurate predictions. A standard GRU operates on data sequentially and has a hidden state that serves as a memory to retain previous information. Nevertheless, it has only acquired data from earlier occurrences. A BiGRU comprises two GRUs operating in opposing directions: one GRU processes the sequence from the beginning to the end, whereas the other GRU processes it from the end to the beginning. These outputs are combined at each time step to obtain the entire sequence by incorporating details from the previous and subsequent contexts.

A GRU cell at time step t computes the following:

$$\text{Update gate } zt = \sigma(Wz \times [ht - 1, xt] + bz), \tag{1}$$

$$\text{Reset gate } rt = \sigma(Wr \times [ht - 1, xt] + br), \tag{2}$$

$$\text{Candidate hidden state } ht = tanh(Wh \cdot [rt \times ht - 1, xt] + bh, \tag{3}$$

$$\text{Final hidden state } ht' = zt \times ht - 1 + (1 - zt) \times ht, \tag{4}$$

where σ denotes the sigmoid activation function, tanh is the hyperbolic tangent function, W and b are the weights and biases, respectively, *xt* is the input at time t, and ht is the hidden state at time t. The BiGRU contains two hidden states at each time step, namely *ht(fwd)* and *ht(bwd)*, calculated by the forward and backward GRUs, respectively. The forward GRU processes the sequence in the conventional manner, whereas the backward GRU processes it in the opposite direction. The combined hidden state at each time step t is acquired by either concatenating or summing the forward and backward hidden states:

$$ht(bi) = ht(fwd) + ht(bwd). \tag{5}$$

The BiGRU can capture dependencies and patterns that may be overlooked by a normal GRU, particularly in sequences wherein the future and past contexts are equally notable. This approach is frequently used in applications such as sequence labeling, time-series prediction, and natural language processing. BiGRUs have a higher computational cost and more parameters than regular GRUs, which may result in overfitting when working with smaller datasets. Thus, the BiGRU enhances the functionality of the regular GRU by including inputs from the forward and backward directions of a sequence, resulting in a more holistic comprehension of the context.

The self-attention mechanism primarily focuses on the internal dependence of an input (Fig. 5). The output of the neural unit in the current moment may probably be affected by the EEG and EOG signals representation in the form of CMW transform. Based on the degree of influence, different weight parameters were assigned to signals such that the model can pay attention to the pivotal signals of stress information. The scaled dot-product attention model was used herein to optimize the data:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (6)$$

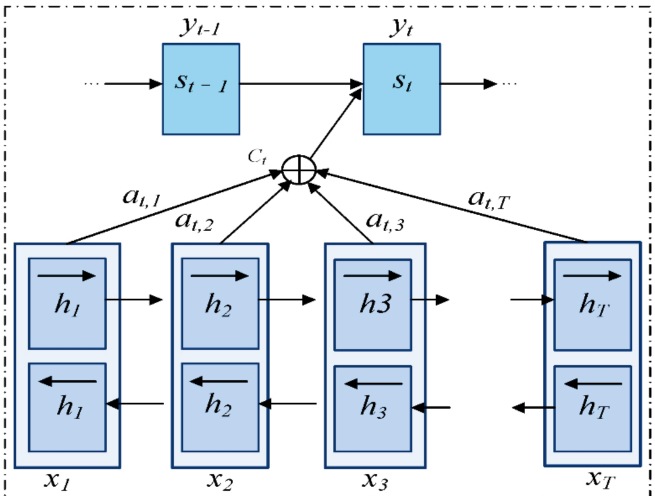


Fig. 5. A self-attention layer integrated into the BiGRU model.

The matrices Q, K, and V comprise query vectors, key vectors, and value vectors, respectively, and dk is the dimension of the input vector. In self-attention mechanism, Q, K, and V were derived from the same input and similarity among all words in the sentence was calculated. The greater the similarity among the signals, the stronger the correlation among them; thus, the dependency within the sentence can be captured. In question classification, the impact of each clause and individual word in the question varies. Certain words or sentences are crucial in question classification, whereas others have less impact. To effectively capture the relevant information in a question, an attention mechanism was added to the BiGRU model. This approach aims to highlight key semantic features, extract useful information, and accurately assess the contribution of each word for the classification of the entire question. By doing so, it ensures that the most crucial information is retained while filtering out redundant information, enhancing the efficiency and performance of question classification. The layer network uses the output of the upper layer network model as the input for the layer model, resulting in the vector representation of each sentence in the BiGRU model with the self-attention mechanism.

The basic form of self-attention mechanism can be expressed as follows:

$$S = \tanh(M)$$

$$\alpha = \text{softmax}(\text{W}_n \times \text{S}) \quad (6)$$

$$r = M \times \alpha_n$$

$$q = \tanh(r)$$

D. Feature Classification

Algorithm 1 summarizes stress classification using the BiGRU model with self-attention mechanism (BiGRU) and ensemble-bagged tree classifier. The machine learning method EBTC [41] blends bagging (bootstrap aggregating) with decision trees. It is especially helpful in minimizing variance, preventing overfitting, and boosting model resilience. The EBTC minimizes prediction variance by averaging many trees, which is useful when individual trees overfit. Bootstrapping randomly creates variation among ensemble trees, essential to the approach's efficacy. The ensemble's averaging effect makes the EBTC resilient to data outliers and noise. Bagged ensemble learning improves machine learning algorithm stability and accuracy. Multiple predictors are utilized to create an aggregated predictor. Classification and regression employ decision trees. They made a decision tree by subdividing the data by feature value testing. From the original dataset, multiple bootstrap samples are randomly chosen subsets of data (with replacement) of the same size. Independent decision trees are trained for each bootstrap sample. Training data is different for each tree owing to random sampling with replacement.

Algorithm 1: Sleep disorder classification using EEG and EOG signals and BiGRU with a self-attention layer and EBTC.	
Input	• EOG_data: Array of EOG signal data • EEG_data: Array of EEG signal data • sampling_rate: Sampling rate of the EOG and EEG data • wavelet_parameters: Parameters for the complex wavelet transform
Output	• Final-classification.
Step 1:	Feature Extraction: Filtered_EOG = bandpass_filter(EOG_data, lowcut, highcut, sampling_rate) Filtered_EEG = bandpass_filter(EEG_data, lowcut, highcut, sampling_rate) • Wavelet_Features_EOG←Morlet-WaveletTransform (Filtered_EOG,wavelet_parameters) • Wavelet_Features_EEG←Morlet-waveletTransform(Filtered_EEG,wavelet_parameters) • Combined_Features = concatenate((Wavelet_Features_EEG, Wavelet_Features_EOG), axis=1) • Combined_Image ← stacking(Combined_Features)
Step 2:	Gated Recurrent Unit (GRU) Classifier: GRU_Model←InitializeGRU(layer_parameters). Hidden_States ← GRU_Model(Combined_Image)
Step 3:	Attention Mechanism: • The BiGRU produces a sequence of hidden states ht. • Attention *scores α t* are computed for each hidden state as stated in Eq. (8).

Step 4:	Weighted Sum: Attention_Scores = compute_attention_scores(Hidden_States) Weighted_Hidden_States ← Attention_Scores @ Hidden_States
Step 5:	Ensemble-bagged Tree Classifier (EBTC): • EBTC_Model ← TrainEBTC(Weighted_Hidden_States, labels) • EBTC_Predictions ← EBTC_Model.predict(Weighted_Hidden_States)
Step 6:	Combination and Final Classification: Final_classification ← classify(EBTC_Predictions)
Step 7:	Evaluation: metrics ← evaluate(Final_classification, true_labels)

After all decision trees are trained, the ensemble model makes predictions by aggregating the predictions from all individual trees. This approach is followed for classification tasks via majority voting, wherein each tree votes for a class, and the class with the most votes is considered the ensemble's prediction.

A dataset D with N instances is considered for the EBTC with M trees. For each tree m = 1,2, ..., M), a bootstrap sample Dm is created by randomly selecting N instances from D with replacement. A decision tree Tm is then trained on Dm. For a new instance x, the prediction y is given by Eq. (7).

$$y = model\{T1(x), T2(x), \ldots, TM(x)\}. \tag{7}$$

The EBTC minimizes prediction variance by averaging many trees, which is useful if some trees overfit. The attention mechanism calculates weights using a straightforward scoring function such as a dot product and a Softmax. For a hidden state ht, the attention score αt is calculated as follows:

$$y\alpha t = \sum_T^i exp\big(score(hi)\big)exp\big(score(ht)\big) \tag{8}$$

where score(ht) is a dot product of the hidden state ht and some learnable parameter and T is the length of the sequence. The model then computes a weighted sum of the hidden states using attention weights, thereby generating a context vector that encapsulates the most relevant information from the entire sequence.

$$Y_c^{context-Vector} = \sum_T^i \alpha t \cdot ht \tag{9}$$

A fully connected layer then uses this context vector for the final classification (such as determining the type of sleep disorder). When analyzing EOG and EEG signal analysis for sleep disorders, the attention mechanism assigns higher weights to hidden states corresponding to signal patterns characteristic of certain sleep disorders. In contrast, the signal's ordinary or less informative patterns are assigned lower weights. The BiGRU model prioritizes essential sections of the EOG and EEG signals with a higher predictive value for various sleep disorders, thereby improving the classification accuracy and efficiency.

IV. Experimental Results

This section overviews the creation of distinct subsets of data and classification performed to identify various sleep disorders. The integrated deep learning and machine learning models, including the meta-learner, were trained and tested on a highly advanced computational infrastructure comprising a notebook with powerful specifications to manage the computing requirements of the models effectively. The processor was an Intel(R) Core (TM) i7, 10th Generation, with an essential clock speed of 3.34 GHz, outfitted with four cores and eight logical processors for efficient parallel computing. The machine was equipped with 32GB RAM to effectively cater to the demanding memory requirements of training and testing deep learning models. Windows 10 offers a reliable and compatible platform for various machine-learning operations. The deep learning models were constructed and trained using the TensorFlow framework in conjunction with the Keras framework.

A. *Performance Measures*

The categorization performances of deep learning and machine learning models were compared using traditional measurement metrics. These measures are crucial for fully grasping the effectiveness of the models in terms of different aspects of categorization performance. The metrics and their corresponding calculation algorithms are outlined below:

1) Accuracy: This metric represents the proportion of true results (true positives and negatives) among the total number of cases examined.

$$Accuracy\ (ACC) = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

2) Sensitivity (SE) or Recall (RC): It assesses the model's false negative avoidance.

$$Recall(RC) = Senstivity\ (SE) = \frac{TP}{TP+FN} \tag{11}$$

3) Specificity (SP): This metric assesses the proportion of actual negatives that are correctly identified.

$$Specificity\ (SP) = \frac{TN}{TN+FP} \tag{12}$$

4) F1-score (FS): This metric is a harmonic mean of precision and recall and balances the two.

$$F1 - score\ (FS) = 2 \times \frac{PR \times RC}{PR+RC} \tag{13}$$

In this Eq. (13), the RC and PR are precision and recall metrics calculated from Eq. (11) to Eq. (14), respectively.

$$Precision\ (PR) = \frac{TP}{TP+FP} \tag{14}$$

In addition, a confusion matrix is used to assess the efficacy of the classification models.

- TN: This refers to accurately predicted negative occurrences.
- FP: This refers to incorrectly predicted positive observations.
- FN: This refers to incorrectly predicted negative observations.
- TP: This refers to accurately predicted positive observations.

These indicators assessed the performance of classification models, emphasizing their strengths and identifying areas for enhancement. This provided valuable information regarding the model's ability to accurately differentiate between several classes and maintain a balance between precision and recall. In addition, the area under the receiver operating characteristic curve (AUC) is also utilized to measure the performance of various systems.

B. Experimental Results

The classification performance of models in differentiating healthy subjects from those with sleep disorders and identifying specific sleep disorders is discussed as follows. Tables IV–VI present the results obtained using the proposed system. Several experiments are performed to identify different classes of sleep disorders compared to normal stage. The required hyperparameters are described in Table III. These parameters are defined based on the several experiments.

TABLE III. HYPERPARAMETERS SETUP REQUIRED BY PROPOSED SYSTEM

Models	Hyperparameter	Values
BiGRU	Units	128
	Layers	2
	Dropout Rate	0.3
	Recurrent Dropout Rate	0.2
	Activation Function	tanh
	Batch Size	64
	Learning Rate	0.001
Self-Attention Layer	Attention Size	Equal to Number of Units
	Attention Activation	Softmax scores
EBTC	Number of Estimators	200
	Maximum Depth of Trees	30
	Minimum Samples Split	5
	Minimum Samples Leaf	2
	Bootstrap Samples	True
	Max Features	sqrt

This research, as presented in Table IV, has yielded significant classification results. We have compared the performance of a proposed approach that utilizes complex-morelet wavelet (CMW) decomposition with sleep stages, against a method that uses discrete wavelet transform (DWT) to convert 1D sleep signals into a combined 2D-CMW image. The choice between DWT and CMW is crucial and depends on the needs of the particular analysis. While DWT may be a better choice for computationally efficient broad feature extraction, the CMW generates a scalogram, a 2D array that provides a detailed view of the signal's frequency content over time. This detailed view is crucial for deep analysis that requires frequency and phase information. By converting 1D sleep problem signals into 2D, CMW may provide more significant information for diagnosing and comprehending complicated sleep events.

The advantage of the proposed approach is evident in significantly improved accuracy and area under the curve (AUC) across all sleep disorders. In this context, 'accuracy' refers to the percentage of correctly classified sleep stages or disorders, while 'AUC' is a measure of the model's ability to distinguish between different sleep stages or disorders. For instance, accuracy increased in the case of insomnia from 81.45% with DWT to an impressive 99.70% with CMW, and AUC rose from 0.822 to 0.997. Similar enhancements are observed across other disorders, such as narcolepsy; accuracy improved from 80.10% to 97.60%, and AUC increased from 0.834 to 0.976. In the case of NFLE, accuracy rose from 82.47% to 95.40%, with AUC increasing from 0.833 to 0.954. In the case of PLM, there was an increase in accuracy from 85.67% to 94.50% and in AUC from 0.865 to 0.945. For RBD, accuracy increased from 84.32% to 96.50%, with AUC improving from 0.854 to 0.965. For SDB, accuracy jumped from 85.20% to 98.30%, and AUC from 0.876 to 0.983. For healthy individuals, accuracy went from 87.00% to 94.10%, with the AUC moving from 0.876 to 0.941. Accordingly, our research has demonstrated the remarkable improvements in accuracy and AUC that the proposed approach offers compared to the method using discrete wavelet decomposition. These improvements underscore the potential of our approach to enhance the diagnosis and treatment of sleep disorders. These findings unequivocally demonstrate the value of incorporating CMW into the classification procedure for a more precise and trustworthy diagnosis of sleep disorders.

TABLE IV. CLASSIFICATION RESULTS OBTAINED WITHOUT USING WAVELET DECOMPOSITION WITH THE SLEEP STAGES. RESULTS ARE OBTAINED USING 10-FOLD CROSS-VALIDATION

Disorder	Discrete Wavelet Transform (DWT)		Complex-Morlet-wavelet (CMW) transform	
	Acc (%)	AUC	Acc (%)	AUC
Insomnia	81.45	0.822	99.70	0.997
Narcolepsy	80.10	0.834	97.60	0.976
NFLE	82.47	0.833	95.40	0.954
PLM	85.67	0.865	94.50	0.945
RBD	84.32	0.854	96.50	0.965
SDB	85.20	0.876	98.30	0.983
Healthy	87.00	0.876	94.10	0.941
Average	83.74	0.858	96.59	0.966

Table V presents sleep disorder classification results obtained using various classifiers with the hold-out validation strategy. The proposed approach, CMW-BiGRU-Self-attention-EBTC, demonstrates superior performance across all sleep disorders compared to alternative methods such as LSTM and GRU-SVM. The proposed CMW-BiGRU-Self-attention-EBTC method shows significant advantages over alternative approaches, highlighting its effectiveness in accurately classifying sleep disorders.

An independent experiment was conducted to compare the CMW transform with spectrogram and scalogram techniques. In practice, the CMW transforms provide both amplitude and phase information, which are essential for detailed signal analysis, particularly in identifying phase coupling between different signal components. The CMWT is particularly advantageous due to its rich and detailed time-frequency representation. Table VI offers a comprehensive overview of such classification performance for a six-class sleep disorder classification task with signal transformation using a 2D-spectogram image. Fig. 6 demonstrates the various confusion matrices, illustrating the distribution of actual and predicted classes, and the number of instances correctly and incorrectly classified for each sleep disorder when a 2D-scalogram is used. The results in Table VI were obtained without using CMW transforms, but the 2D-spectogram was used to analyze the performance. As shown in this table, the proposed system, which does not incorporate the complex morlet wavelet (CMW) transformation, the bi-directional gated recurrent unit (BiGRU) with self-attention, and the ensemble bagged tree classifier (EBTC) is utilized. The result does not outperform by the system without using CMW transforms. Similar trends were observed using 2D-scalogram without using the CMW transform technique, as depicted in Fig. 6. Hence, the proposed CMW-BiGRU-Self-attention-EBTC system using CMW transform ensures accurate sleep disorder classification. Despite these results, the CMW transform's adaptability and detailed time-frequency representation continue to outperform spectrograms and scalograms, particularly in capturing complex, non-stationary signal patterns essential for diagnosing sleep disorders.

TABLE V. SLEEP DISORDER CLASSIFICATION RESULTS FROM USING VARIOUS MACHINE-LEARNING CLASSIFIERS WITH THE HOLD-OUT VALIDATION STRATEGY

Classifier	Accuracy (%)						
	Insomnia	Narcolepsy	NFLE	PLM	RBD	SDB	Healthy
CMW-BiGRU-Self-attention-EBTC	99.70	97.60	95.40	94.50	96.50	98.30	94.10
LSTM	80.9	75.3	74.5	77.1	79.8	81.0	80.9
KNN+NN	78.6	73.5	71.9	74.8	77.4	79.2	78.6
Random Forest (RF)	92.5	87.0	85.1	88.2	90.3	91.2	92.5
RF+SVM	90.8	85.5	83.3	87.7	89.4	90.1	90.8
GRU-SVM	89.7	84.2	82.9	86.5	88.0	89.3	89.7
Decision Tree	85.4	80.6	79.2	81.9	83.7	85.1	85.4
AdaBoost	83.2	78.4	76.8	80.0	82.0	83.5	83.2

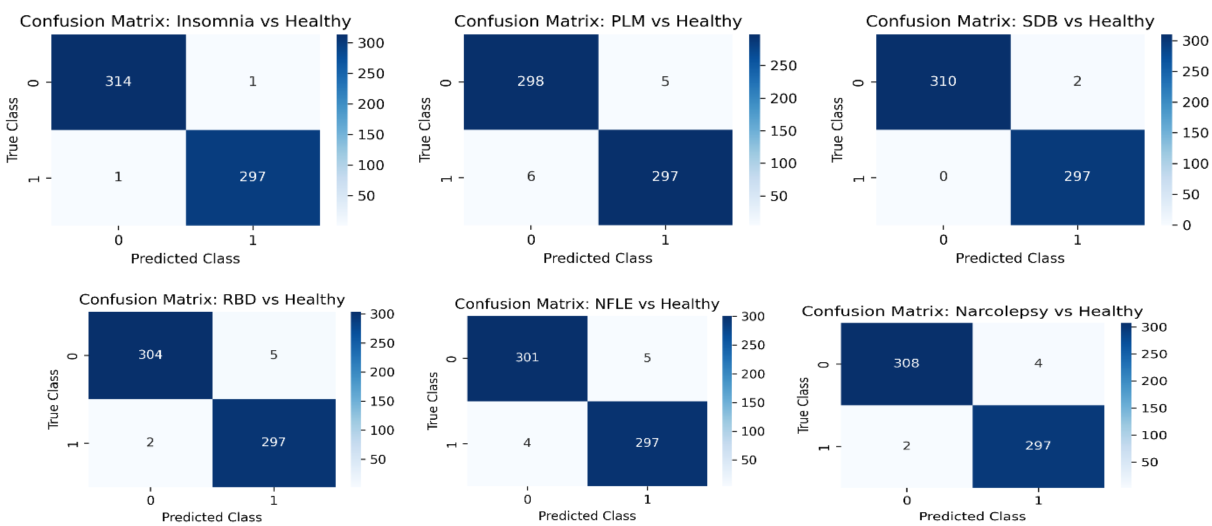


Fig. 6. Confusion metrics obtained for six sleep disorder prediction versus healthy individuals without signal transformation (normalization, bandpass filter and used 2D-scalogram image, complex morlet wavelet, stacking) step by proposed CMW-BiGRU-Self-attention-EBTC system.

TABLE VI. PERFORMANCE METRICS OBTAINED FOR SIX-CLASS BASED SLEEP DISORDER CLASSIFICATION WITHOUT SIGNAL TRANSFORMATION (NORMALIZATION, BANDPASS FILTER AND USED 2D-SPECTOGRAM IMAGE, COMPLEX MORLET WAVELET, STACKING) STEP BY PROPOSED CMW-BIGRU-SELF-ATTENTION-EBTC SYSTEM

Classes	ACC (%)	PR(%)	RC(%)	FS(%)
Insomnia	84.75	84.77	84.77	84.77
Narcolepsy	82.96	82.93	82.93	82.93
NFLE	81.09	81.11	81.11	81.11
PLM	80.33	80.42	80.42	80.42
RBD	82.03	82.03	82.03	82.03
SDB	83.56	83.62	83.62	83.62
Healthy	79.99	79.96	79.96	79.96

This study distinguishes healthy people from individuals with sleep problems with great accuracy, as described in the above paragraphs. Several variables were found and studied throughout testing, which are potential sources of error or misclassification based on the experimental results from Tables IV and V, as described below.

- Noise from EEG and EOG signal quality can decrease feature extraction precision. Signal fluctuation can be caused by electrode location, sleep movement, and physiological variations. When the signal-to-noise ratio is low, this fluctuation might cause the model to categorize epochs inconsistently.
- Despite the Use of cross-validation and regularization to reduce overfitting, the model may have inadvertently

learned patterns specific to the training set that do not generalize well to unknown data. This is a common challenge in machine learning that could potentially increase model error rates when confronted with new data.

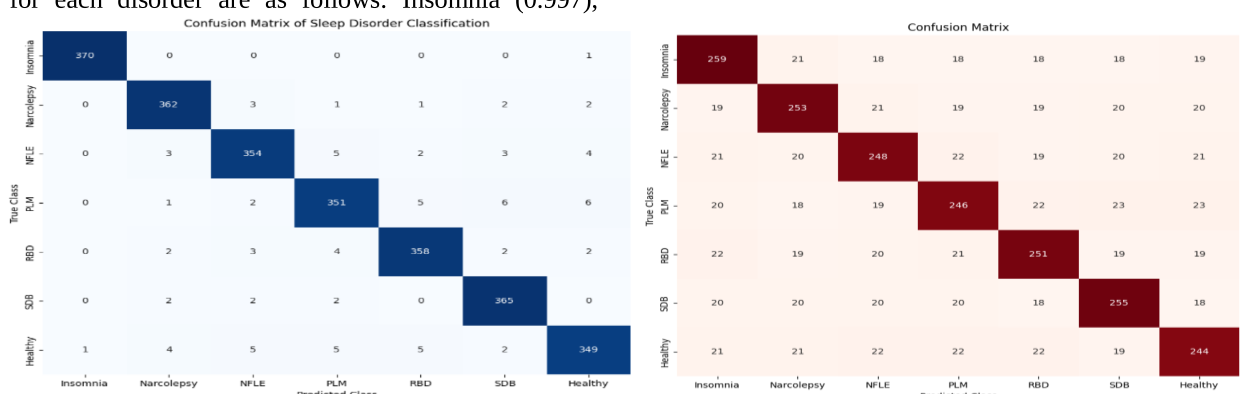
- This dataset is extensive; however, class imbalance is evident in SDB and PLM data. Class imbalance can bias the model to recognize the dominant class but not the minority class, reducing accuracy and increasing misclassification rates.
- Rare sleep disorders, including Narcolepsy and Nocturnal Frontal Lobe Epilepsy (NFLE), reduce training data. This constraint might hinder the model's understanding of these illnesses' complicated patterns, leading to reduced accuracy or increased misclassification rates.
- The proposed model uses Morlet wavelet transform and advanced machine learning methods to analyze combine EEG and EOG signals, requiring numerous computing layers. Complexity allows excellent accuracy, but it also raises the possibility of misclassifying signal artifacts or non-standard signal patterns as disorders.

To address these potential errors, we implemented a series of robust measures. We strengthened the signal processing capabilities with CMW transforms. Additionally, training dataset was expanded to encompass a wider range of patients and disorders.

Fig. 7–8 present the results obtained using the proposed system by changing the hyperparameters to detect six multi-class sleep stages. ReLU activation function is applied on GRU units as compared to Tanh as shown in Fig. 7. A similar performance has been achieved. This figure presents the receiver operating characteristic (ROC) curves for different sleep disorders, as well as for the 'Healthy' class, based on the crucial area under the curve (AUC) values provided. Each curve represents the trade-off between the actual positive rate (sensitivity) and the false positive rate (1-specificity) for a specific disorder classification. A higher AUC indicates better discrimination ability, with values closer to 1.0 representing superior classification performance. In this figure, the AUC values for each disorder are as follows: Insomnia (0.997), Narcolepsy (0.976), NFLE (0.954), PLM (0.945), RBD (0.965), SDB (0.983), and Healthy (0.941). These values are pivotal in reflecting the models' ability to distinguish between positive instances of each disorder and negative instances of other disorders or healthy individuals. AUC values near 1.0 suggest excellent classification performance, while lower values indicate room for improvement.

The confusion matrix in Fig. 8 provides a comprehensive overview of a CNN classifier's performance in discerning between various sleep disorders and healthy individuals. Each row corresponds to the true class, while each column represents the predicted class, offering insights into both correct classifications and misclassifications when figure (a) proposed system and (b) original BiLSTM and EBTC boosting tree. For instance, the classifier demonstrated its effectiveness by accurately identifying a significant number of instances for each disorder, such as Insomnia (370 instances correctly classified) and Narcolepsy (362 instances correctly classified). However, misclassifications were observed across different categories, indicating the model's limitations in certain scenarios. Notably, while the classifier performed well in identifying instances of Insomnia and Narcolepsy, a few instances of Insomnia (1 instance misclassified as Healthy) and Narcolepsy (2 instances misclassified as NFLE, one as PLM, one as RBD, two as SDB, and two as Healthy) were incorrectly classified.

Moreover, the confusion matrix highlighted misclassifications of healthy individuals, with some instances erroneously classified as various sleep disorders. Despite accurately identifying the majority of healthy individuals (349 instances correctly classified), a notable number of misclassifications occurred. For example, healthy individuals were mistakenly classified as Narcolepsy (4 instances misclassified), NFLE (5 instances misclassified), PLM (5 instances misclassified), RBD (5 instances misclassified), and SDB (2 instances misclassified). These misclassifications underscore the need to refine the classification model to enhance its accuracy and robustness, particularly distinguishing between healthy individuals and those with sleep disorders. By addressing these misclassifications and improving the classifier's ability to identify different sleep patterns accurately, it can strengthen diagnostic tools for sleep disorders and enhance patient care through more precise and timely interventions.



(a) (b)

Fig. 7. Confusion matrices for sleep stage and sleep disorder detection using (a) proposed system and (b) original BiLSTM and EBTC boosting tree.

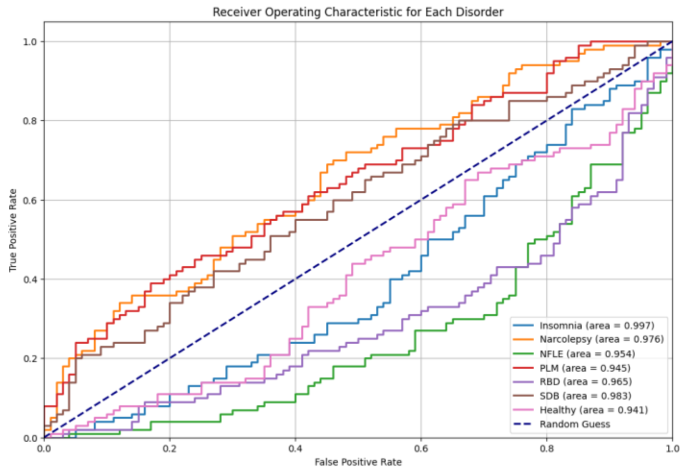


Fig. 8. A separate experiment for using 2400 samples of each sleep-disorder class to show AUC curves of accuracy with respect to each class of sleep disorder.

The AUC curve for our proposed system using the architecture (CMW-BiGRU-Self-attention-EBTC) is shown in Fig. 8. This is an integral part of our research. This system, which uses a large dataset with 24,000 samples of each class instead of 26,000 samples, gives a full picture of how well the proposed classifier works at finding sleep disorders. More importantly, it paves the way for evaluating and refining the classification model, thereby enhancing the accuracy and reliability of our system. This, in turn, underscores the potential impact of our research on clinical applications, making it a significant contribution to the fields of sleep medicine and machine learning.

C. State-of-the-Art Comparisions

Sleep disorder diagnosis through automated systems has seen significant advancements, with various models employed in the past by combinations of convolutional neural networks (CNNs), recurrent neural networks (RNNs), gated recurrent units (GRUs), and random forest (RF) algorithms. These models, such as Shao-CNN-GRU [17], Fan-CNN-RNN [19], Santaji-RF [21], Morokuma-CNN [30], and Satapathy-1DCNN [33], have leveraged electrooculography (EOG) and electroencephalography (EEG) signals to varying degrees of success. These systems were selected to perform State-of-the-art (SOTA) comparisons because they were quickly implemented. While they have shown promising results, challenges such as overfitting, limited sleep class detection, and computational inefficiency have persisted, as mentioned in Table I. Accuracies have varied widely, from as low as 76.3% to as high as 98.84%, with each model demonstrating its strengths in handling complex signal data for sleep stage classification or disorder diagnosis and revealing significant limitations hindering their broader application and effectiveness.

In this context, the proposed CMW-BiGRUSelf-EBTC model emerges as a noteworthy evolution that adeptly addresses these challenges. By integrating a complex metro-wavelet transformation with a bidirectional gated recurrent unit that includes a self-attention layer and coupling it with an ensemble-bagged tree classifier, this model simplifies the signal processing pipeline and enhances the accuracy and efficiency of sleep disorder diagnosis. With remarkable success rates across various sleep disorders and an overall classification accuracy of 96% alongside an AUC of 0.96, this approach outshines its predecessors. It achieves this by effectively mitigating overfitting and reducing computational demands, thereby marking a significant leap forward in developing noninvasive, automated systems for accurate and efficient sleep disorder diagnosis. Specific hyperparameters are described in Table VII to expand the "Details" column to include examples of hyperparameters that might be tuned for RNNs (Recurrent Neural Networks) and the considerations for architectural configurations.

Table VIII shows a comparative analysis of various state-of-the-art (SOTA) systems for sleep disorder detection. The systems compared include Shao-CNN-GRU, Fan-CNN-RNN, Santaji-RF, Morokuma-CNN, Satapathy-1DCNN, and the proposed system, CMW-BiGRUSelf-EBTC.

TABLE VII. STATE-OF-THE-ART COMPARISONS (SOTA) HYPER-PARAMETERS SITTING

Component	Description	Expected Impact
Baseline (Full Model)	CMW-BiGRUSelf-EBTC with all features	Establish baseline performance
Without Bidirectional RNN Cells	Use unidirectional GRU cells	Assess the importance of capturing temporal dependencies in both directions
Without the Attention Layer	Remove the self-attention layer	Evaluate the impact of focusing on significant parts of the data
Without Bagging Ensemble	Use a single classifier instead of an ensemble	Determine the contribution of ensemble methods to robustness and accuracy
Without Data Augmentation	Train without augmented data	Examine the role of data diversity in model generalization

TABLE VIII. SOTA WITH RESPECT TO PROPOSED SYSTEM

SOTA Systems	Acc (%)	SE(%)	SP(%)	AUC	FS (%)
Shao-CNN-GRU [17]	88.25	86.40	88.15	0.87	88.32
Fan-CNN-RNN [19]	78.44	77.12	78.42	0.77	78.46
Santaji-RF [21]	75.47	73.70	74.10	0.75	76.10
Morokuma-CNNc[30]	86.50	85.20	88.45	0.86	88.00
Satapathy-1DCNN [33]	80.10	82.66	83.25	0.83	82.12
CMW-BiGRUSelf-EBTC	96.59	97.30	95.20	0.966	96.0

TABLE IX. ABLATION STUDY PARAMETERS

Step	Description	Details	Example Hyperparameters
Define the Parameter Space	For each model, identify and list all hyperparameters and architectural configurations that will be explored.	Includes learning rates, batch sizes, number of layers, types of RNN cells (e.g., LSTM, GRU), etc.	Learning rate: [0.001], Batch size: [64], RNN type: [LSTM, GRU], Number of layers: [3], Dropout rate: [0.25]
Apply Nested CV	Implement nested cross-validation (CV) with an outer loop for performance assessment and an inner loop for hyperparameter tuning.	Ensures unbiased evaluation and that hyperparameter tuning does not influence the test set.	Outer loop: 5 folds, Inner loop: 3 folds
Optimize Hyperparameters	Use grid search within the inner loop of the nested CV to find the optimal set of hyperparameters and architectural configurations for each model.	Systematically explores multiple combinations of parameters to find the best setup for each model.	Grid search across all combinations of the example hyperparameters listed.
Evaluate and compare	After tuning, evaluate each model's performance on the test set of the outer CV loop to ensure fairness in comparison.	Performance metrics are based on unseen data, providing a reliable basis for comparison.	Accuracy, F1 Score, AUC, Sensitivity, Specificity
Statistical Testing	Employ statistical tests to determine if the differences in performance between models are statistically significant.	Strengthens the validity of the comparison by confirming whether observed performance differences are meaningful.	Wilcoxon signed-rank tests comparing model performances.

The recommended CMW-BiGRUSelf-EBTC system, with its superior performance metrics, outperforms previous models. Its accuracy of 96.59%, SE of 97.30%, SP of 95.20%, AUC of 0.966, and F1-score of 96.0% are a testament to its effectiveness. The system's ability to effectively diagnose sleep problems is a significant advancement, demonstrating its profound impact on sleep problem detection. This comparison underscores the importance of advanced machine learning approaches like bidirectional RNNs, self-attention layers, and ensemble methods in the field of sleep disorder diagnosis.

D. Ablation Study

Ablation studies on the proposed Complex-Morlet-wavelet Representation using a bidirectional gated recurrent unit with a Self-attention Layer and an ensemble-bagged tree classifier (CMW-BiGRUSelf-EBTC) system involve systematically removing or replacing model components to understand their performance contributions. This study highlights the most critical elements of the suggested sleep problem detection and identification technique—ablation research structure. Tables VIII and IX provide an overview of the CMW-BiGRUSelf-EBTC system ablation research and a full analysis of each ablation component.

These tables summarize the ablation study's setup and findings. They also demonstrate how each CMW-BiGRUSelf-EBTC component detected and diagnosed numerous sleep problems. The ablation research for the planned CMW-BiGRUSelf-EBTC system employing EEG and EOG data to discover and diagnose sleep problems shows how the elements work together to make it operate successfully. Initial evaluation of the system yields remarkable metrics: accuracy of 96.59%, sensitivity of 97.30%, specificity of 95.20%, AUC and F-score of 0.966 and 96.0%, respectively. This complete performance shows the system's resilience and accuracy in diagnosing sleep problems.

As the study progresses through its phases, removing critical system features one by one, a clear picture of their contributions emerges. All metrics go down a lot when there are no bidirectional RNN cells. This shows the importance of capturing temporal dependencies in the signal data for correct disorder recognition. In the same way, getting rid of the attention layer lowers performance metrics, showing how important it is for helping the model focus on essential parts of the complex signal data. The system's performance worsens when the bagging ensemble method and data augmentation are removed. This shows how important they were for making the model more robust and able to generalize across different data representations. Each component's removal delineates a stepwise decrease in the system's effectiveness, underlining the synergistic effect of these elements in achieving the CMW-BiGRUSelf-EBTC system's state-of-the-art performance.

V. CONCLUSION

The EOG and EEG data were used for automated sleep-disorder identification, making this study a notable advancement in sleep medicine. The proposed method for detecting sleep disorders yielded highly accurate and efficient results as it integrated advanced signal-processing techniques with powerful machine-learning models. This approach was also designed to be patient friendly. Thus, this study not only enhances the scientific comprehension of sleep health but also holds the potential to considerably improve the quality of life of individuals with sleep disorders. The proposed method effectively classifies healthy individuals from those with different sleep disorders. It achieved a remarkably high accuracy of 99.7% for insomnia, 97.6% for narcolepsy, 95.4% for NFLE, 94.5% for PLM, 96.5% for RBD, 98.3% for SDB, and 94.1% for healthy individuals. The model's relevance and precision can be enhanced across many scenarios by establishing a confidential database for subsequent experimentation.

A key priority is ensuring the wide-ranging appropriateness and efficacy of a proposed technique for sleep health monitoring and diagnosis across diverse demographic groups. The dataset was expanded to include additional ages, genders, ethnicities, and geographical origins to achieve this. This expanded dataset will better capture population-specific sleep patterns and problems, enhancing the model's generalizability. To understand how cultural influences impact sleep, it will employ cross-cultural validation and adaptive algorithms for tailored diagnosis. However, medical specialists from diverse demographics must refine the model to maintain its clinical relevance and responsiveness to the vast range of sleep problems.

Ethical and inclusive research and design practices emphasize privacy, permission, and data protection. By making this technology inexpensive and accessible across demographics, it promotes healthcare equity. Patients and healthcare providers must monitor and offer feedback post-deployment. The system will be modified and adjusted based on real-world use and feedback to keep it valuable and relevant for diagnosing and monitoring sleep problems in varied worldwide populations.

Data Availability Statement: This study used the PhysioNet CAP Sleep database of the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy, as downloaded via physionet.org from https://physionet.org/content/capslpdb/1.0.0/ (accessed on March 5, 2022).

ACKNOWLEDGMENT

The PhysioNet CAP Sleep database of the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy, was used via physionet.org.

REFERENCES

[1] Xu, S.; Faust, O.; Seoni, S.; Chakraborty, S.; Barua, P.D.; Loh, H.W.; Elphick, H.; Molinari, F.; Acharya, U.R. A review of automated sleep disorder detection. Comput. Biol. Med. 2022, 150, 106100. DOI:10.1016/j.compbiomed.2022.106100.

[2] Loh, H.W.; Ooi, C.P.; Vicnesh, J.; Oh, S.L.; Faust, O.; Gertych, A.; Acharya, U.R. Automated detection of sleep stages using deep learning techniques: A systematic review of the last decade (2010–2020). Appl. Sci. 2020, 10, 8963. DOI:10.3390/app10248963.

[3] Rim, B.; Sung, N.J.; Min, S.; Hong, M. Deep learning in physiological signal data: A survey. Sensors (Basel). 2020, 20, 969. DOI:10.3390/s20040969.

[4] Van Der Donckt, J.; Van Der Donckt, J.; Deprost, E.; Vandenbussche, N.; Rademaker, M.; Vandewiele, G.; Van Hoecke, S. Do not sleep on traditional machine learning. Biomed. Signal Process. Control. 2023, 81, 104429. DOI:10.1016/j.bspc.2022.104429.

[5] Fatimah, B.; Singhal, A.; Singh, P. A multi-modal assessment of sleep stages using adaptive Fourier decomposition and machine learning. Comput. Biol. Med. 2022, 148, 105877. DOI:10.1016/j.compbiomed.2022.105877.

[6] Faust, O.; Hagiwara, Y.; Hong, T.J.; Lih, O.S.; Acharya, U.R. Deep learning for healthcare applications based on physiological signals: a review. Comput. Methods Programs Biomed. 2018, 161, 1–13. DOI:10.1016/j.cmpb.2018.04.005.

[7] Faust, O.; Razaghi, H.; Barika, R.; Ciaccio, E.J.; Acharya, U.R. A review of automated sleep stage scoring based on physiological signals for the new millennia. Comput. Methods Programs Biomed. 2019, 176, 81–91. DOI:10.1016/j.cmpb.2019.04.032.

[8] Aboalayon, K.A.I.; Faezipour, M.; Almuhammadi, W.S.; Moslehpour, S. Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation. Entropy. 2016, 18, 272. DOI:10.3390/e18090272.

[9] Qian, X.; Qiu, Y.; He, Q.; Lu, Y.; Lin, H.; Xu, F.; Zhu, F.; Liu, Z.; Li, X.; Cao, Y.; Shuai, J. A review of methods for sleep arousal detection using polysomnographic signals. Brain Sci. 2021, 11, 1274. DOI:10.3390/brainsci11101274.

[10] Moridian, P.; Shoeibi, A.; Khodatars, M.; Jafari, M.; Pachori, R.B.; Khadem, A.; Alizadehsani, R.; Ling, S.H. Automatic diagnosis of sleep apnea from biomedical signals using artificial intelligence techniques: methods, challenges, and future works. WIREs Data Min. & Knowl. 2022, 12, e1478. DOI:10.1002/widm.1478.

[11] Sharma, M.; Yadav, A.; Tiwari, J.; Karabatak, M.; Yildirim, O.; Acharya, U.R. An automated wavelet-based sleep scoring model using eeg, emg, and eog signals with more than 8000 subjects. Int. J. Environ. Res. Public Health. 2022, 19, 7176. DOI:10.3390/ijerph19127176.

[12] Jarchi, D.; Andreu-Perez, J.; Kiani, M.; Vysata, O.; Kuchynka, J.; Prochazka, A.; Sanei, S. Recognition of patient groups with sleep related disorders using bio-signal processing and deep learning. Sensors (Basel). 2020, 20, 2594. DOI:10.3390/s20092594.

[13] Sharma, M.; Darji, J.; Thakrar, M.; Acharya, U.R. Automated identification of sleep disorders using wavelet-based features extracted from electrooculogram and electromyogram signals. Comput. Biol. Med. 2022, 143, 105224. DOI:10.1016/j.compbiomed.2022.105224.

[14] Sekkal, R.N.; Bereksi-Reguig, F.; Ruiz-Fernandez, D.; Dib, N.; Sekkal, S. Automatic sleep stage classification: From classical machine learning methods to deep learning. Biomed. Signal Process. Control. 2022, 77, 103751. DOI:10.1016/j.bspc.2022.103751.

[15] Sharma, M.; Tiwari, J.; Patel, V.; Acharya, U.R. Automated identification of sleep disorder types using triplet half-band filter and ensemble machine learning techniques with eeg signals. Electronics. 2021, 10, 1531. DOI:10.3390/electronics10131531.

[16] Rahman, M.M.; Bhuiyan, M.I.H.; Hassan, A.R. Sleep stage classification using single-channel EOG. Comput. Biol. Med. 2018, 102, 211–220. DOI:10.1016/j.compbiomed.2018.08.022.

[17] Shao, X.; Soo Kim, C. A hybrid deep learning scheme for multi-channel sleep stage classification. Comput. Mater. Continua. 2022, 71, 889–905. DOI:10.32604/cmc.2022.021830.

[18] Satapathy, S.K.; Kondaveeti, HariKishan; Loganathan, D.; Sharathkumar, S. A machine learning model for automated classification of sleep stages using polysomnography signals. In Machine Vision and Augmented Intelligence—Theory and Applications: Select. Proceedings of MAI 2021; Springer: Berlin, 2021, pp. 209–222. DOI:10.1007/978-981-16-5078-9_18.

[19] Fan, J.; Sun, C.; Long, M.; Chen, C.; Chen, W. Eognet: A novel deep learning model for sleep stage classification based on single-channel eog signal. Front. Neurosci. 2021, 15, 573194. DOI:10.3389/fnins.2021.573194.

[20] Chambon, S.; Galtier, M.N.; Arnal, P.J.; Wainrib, G.; Gramfort, A. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. IEEE Trans. Neural Syst. Rehabil. Eng. 2018, 26, 758–769. DOI:10.1109/TNSRE.2018.2813138.

[21] Santaji, S.; Desai, V. Analysis of EEG signal to classify sleep stages using machine learning. Sleep Vigilance. 2020, 4, 145–152. DOI:10.1007/s41782-020-00101-9.

[22] Mousavi, S.; Afghah, F.; Acharya, U.R. SleepEEGNet: automated sleep stage scoring with sequence to sequence deep learning approach. PLOS ONE. 2019, 14, e0216456. DOI:10.1371/journal.pone.0216456.

[23] Somaskandhan, P.; Leppänen, T.; Terrill, P.I.; Sigurdardottir, S.; Arnardottir, E.S.; Ólafsdóttir, K.A.; Serwatko, M.; Sigurðardóttir, S.Þ.; Clausen, M.; Töyräs, J.; Korkalainen, H. Deep learning-based algorithm accurately classifies sleep stages in preadolescent children with sleep-disordered breathing symptoms and age-matched controls. Front. Neurol. 2023, 14, 1162998. DOI:10.3389/fneur.2023.1162998.

[24] Yan, R.; Li, F.; Zhou, D.D.; Ristaniemi, T.; Cong, F. Automatic sleep scoring: A deep learning architecture for multi-modality time series. J. Neurosci. Methods. 2021, 348, 108971. DOI:10.1016/j.jneumeth.2020.108971.

[25] Satapathy, S.K.; Loganathan, D. Multimodal multiclass machine learning model for automated sleep staging based on time series data. SN Comput. Sci. 2022, 3, 276. DOI:10.1007/s42979-022-01156-3.

[26] Malafeev, A.; Laptev, D.; Bauer, S.; Omlin, X.; Wierzbicka, A.; Wichniak, A.; Jernajczyk, W.; Riener, R.; Buhmann, J.; Achermann, P. Automatic human sleep stage scoring using deep neural networks. Front. Neurosci. 2018, 12, 781. DOI:10.3389/fnins.2018.00781.

[27] Ji, X.; Li, Y.; Wen, P. 3DSleepNet: A multi-channel bio-signal based sleep stages classification method using deep learning. IEEE Trans. Neural Syst. Rehabil. Eng. 2023, 31, 3513–3523. DOI:10.1109/TNSRE.2023.3309542.

[28] Cooray, N.; Andreotti, F.; Lo, C.; Symmonds, M.; Hu, M.T.M.; De Vos, M. Detection of REM sleep behaviour disorder by automated polysomnography analysis. Clin. Neurophysiol. 2019, 130, 505–514. DOI:10.1016/j.clinph.2019.01.011.

[29] Li, Y.; Xu, Z.; Zhang, Y.; Cao, Z.; Chen, H. Automatic sleep stage classification based on a two-channel electrooculogram and one-channel electromyogram. Physiol. Meas. 2022, 43, p.07NT02. DOI:10.1088/1361-6579/ac6bdb.

[30] Morokuma, S.; Hayashi, T.; Kanegae, M.; Mizukami, Y.; Asano, S.; Kimura, I.; Tateizumi, Y.; Ueno, H.; Ikeda, S.; Niizeki, K. Deep learning-based sleep stage classification with cardiorespiratory and body movement activities in individuals with suspected sleep disorders. Sci. Rep. 2023, 13, 17730. DOI:10.1038/s41598-023-45020-7.

[31] Sokolovsky, M.; Guerrero, F.; Paisarnsrisomsuk, S.; Ruiz, C.; Alvarez, S.A. Deep learning for automated feature discovery and classification of sleep stages. IEEE ACM Trans. Comp. Biol. Bioinform. 2020, 17, 1835–1845. DOI:10.1109/TCBB.2019.2912955.

[32] Bozkurt, F.; Uçar, M.K.; Bilgin, C.; Zengin, A. Sleep–wake stage detection with single channel ECG and hybrid machine learning model in patients with obstructive sleep apnea. Phys. Eng. Sci. Med. 2021, 44, 63–77. DOI:10.1007/s13246-020-00953-5.

[33] Satapathy, S.K.; Loganathan, D. Automated classification of multi-class sleep stages classification using polysomnography signals: a nine- layer 1D-convolution neural network approach. Multimedia Tool. Appl. 2023, 82, 8049–8091. DOI:10.1007/s11042-022-13195-2.

[34] Zhao, R.; Xia, Y.; Wang, Q. Dual-modal and multi-scale deep neural networks for sleep staging using EEG and ECG signals. Biomed. Signal Process. Control. 2021, 66, 102455. DOI:10.1016/j.bspc.2021.102455.

[35] Casciola, A.A.; Carlucci, S.K.; Kent, B.A.; Punch, A.M.; Muszynski, M.A.; Zhou, D.; Kazemi, A.; Mirian, M.S.; Valerio, J.; McKeown, M.J.; Nygaard, H.B. A deep learning strategy for automatic sleep staging based on two-channel EEG headband data. Sensors (Basel). 2021, 21, 3316. DOI:10.3390/s21103316.

[36] Jia, Z.; Cai, X.; Zheng, G.; Wang, J.; Lin, Y. SleepPrintNet: A multivariate multimodal neural network based on physiological time-series for automatic sleep staging. IEEE Trans. Artif. Intell. 2020, 1, 248–257. DOI:10.1109/TAI.2021.3060350.

[37] Kemp, B.; Zwinderman, A.H.; Tuk, B.; Kamphuisen, H.A.; Oberyé, J.J.L. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. IEEE Trans. Biomed Eng. 2000, 47, 1185–1194. DOI:10.1109/10.867928.

[38] Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and Physionet: components of a new research resource for complex physiologic signals. Circulation. 2000, 101, E215–E220. DOI:10.1161/01.cir.101.23.e215.

[39] Zhu, Q.; Jiang, X.; Ye, R. Sentiment analysis of review text based on BiGRU-attention and hybrid CNN. IEEE Access. 2021, 9, 149077–149088. DOI:10.1109/ACCESS.2021.3118537.

[40] Zhang, D.; Hong, M.; Zou, L.; Han, F.; He, F.; Tu, Z.; Ren, Y. Attention pooling-based bidirectional gated recurrent units model for sentimental classification. Int. J. Comput. Intell. Syst. 2019, 12, 723–732. DOI:10.2991/ijcis.d.190710.001.

[41] Ensemble methods, bagging, boosting and stacking, URL. Available online: https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205 (Accessed 2 January 2023).