IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 March 2025, accepted 4 April 2025, date of publication 9 April 2025, date of current version 5 May 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3559350

RESEARCH ARTICLE

Research on Sleep Stage Classification of Electroencephalogram Signals Based on CNN and LSTM

LIYING ZHU[1], QINGXIN GUAN[2], YUNQING LIU[1], JIAYUE XU[1], AND SHA MA[1,3,4]
[1]School of Biomedical Engineering, Guangdong Medical University, Dongguan 523808, China
[2]School of Applied Chemistry and Materials, Zhuhai College of Science and Technology, Zhuhai 519040, China
[3]Songshan Lake Innovation Center of Medicine and Engineering, Guangdong Medical University, Dongguan 523808, China
[4]Key Laboratory of Medical Electronics and Medical Imaging Equipment, Dongguan 523808, China

Corresponding author: Sha Ma (sma@gdmu.edu.cn)

This work was supported in part by Guangdong Basic and Applied Basic Foundation Research under Grant 2021A1515110494.

ABSTRACT Chronic sleep deprivation can seriously affect physical and mental health and increase the risk of many chronic diseases. Sleep staging methods can assess sleep quality, but the manual interpretation is inconvenient and subjective. Establishing efficient automated sleep staging models is crucial to saving time and cost, improving efficiency, and helping patients with sleep disorders receive timely treatment. In recent years, artificial intelligence has been rapidly developing in the medical field, diagnostic efficiency has become an auxiliary tool, and even the accuracy rate exceeds that of doctors. In this paper, deep learning technology is applied to sleep EEG signal staging, exploring the basic theory and preprocessing method of EEG signal, and comparing the performance of the automatic sleep staging model through experiments. The process mainly involves processing the original dataset, establishing a standard dataset, and extracting features using a band-pass filter combined with VMD and FFT methods. Two automatic sleep staging models based on LSTM and LSTM+CNN were proposed. The models, which integrate VMD and FFT methods, achieved accuracies of 87% and 92%, respectively. Experimental results showed that the LSTM+CNN model, which combines VMD and FFT methods, performed better in terms of classification accuracy and loss values. Compared to using LSTM alone for sleep staging, it demonstrated outstanding classification performance. This technology is expected to provide healthcare institutions and doctors with faster and more accurate sleep monitoring and diagnostic tools to improve patients' sleep health and quality of life.

INDEX TERMS Automatic sleep stage classification, deep learning, EEG brain signals, variational mode decomposition, LSTM.

I. INTRODUCTION

Sleep plays a vital role in our daily lives. Due to the busyness of daily life, the actual duration of rest for the general public is being shortened. Data suggest that chronic sleep deprivation can seriously affect people's physical and mental health [1]. Sleep staging is an important prerequisite for assessing sleep quality and the first step in addressing sleep disorders [2].

Clinical studies on sleep staging rely on Polysomnography (PSG) recorded by polysomnography monitors. Experienced sleep physicians will manually interpret the PSG signals in 30-second segments based on the Rechtstaffen & Kales (R&K) criteria [3] or the American Academy of Sleep-medicine (AASM) manual [4] to distinguish the current sleep stage and complete the corresponding sleep staging.

While manual interpretation is regarded as the 'gold standard' for sleep staging, it does have certain limitations. Patients need to collect data in a specific sleep monitoring room, which is inconvenient, and sleep staging based on manual interpretation is easily influenced by personal experience and is subjective. Therefore, establishing an efficient and accurate automatic sleep staging model is crucial. The implementation of automatic sleep staging can save time and costs, enhance the efficiency of sleep quality

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague.

assessment, and ensure that patients with sleep disorders receive timely diagnosis and treatment.

In 1968, Rechtschaffen and Kales [3] standardized the staging of the main physiological point signals during sleep, i.e., the R&K sleep staging criteria. As shown in Figure 1, AASM sleep staging classifies sleep into wakefulness(W), Non-rapid eye movement stage 1 (N1), Non-rapid eye movement stage 2(N2), Non-rapid eye movement stage 3(N3) and Rapid Eye Movement(REM) stages based on electroencephalogram (EEG) signals [4].

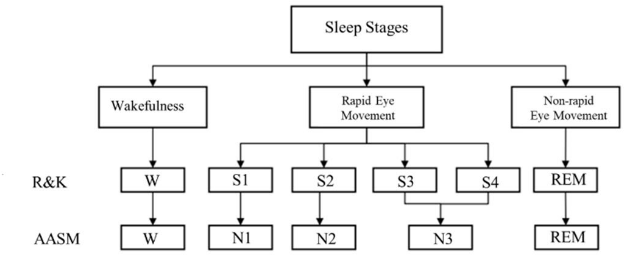

FIGURE 1. R&K and AASM staging criteria.

Automated sleep staging methods in recent years have involved machine learning and deep learning. This paper aims to establish an efficient and accurate automatic sleep staging model to improve the efficiency of sleep quality assessment and provide timely diagnosis and treatment for patients with sleep disorders.

This paper discusses the automatic sleep staging model based on LSTM. Recently, a combined model based on LSTM and Convolutional Neural Networks (CNN) has been investigated.

In the analysis of the experimental results, the model effects of three different time-frequency analysis methods based on LSTM and LSTM+CNN, provide a reference for further optimizing the model structure and algorithm. The experimental results show that the proposed model has potential clinical applications.

II. METHODS

A. DATA ACQUISITION

In this paper, we selected the Sleep-EDF dataset from the Bioinformatics Database of Beth Israel Hospital (MIT-BIH) of Massachusetts Institute of Technology (MIT) [5], a PhysioNet resource-sharing website, which contains two days of PSG data (including physiological signals such as electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG), etc.) from 197 subjects. This dataset contains two days of PSG data (including physiological signals such as EEG, EOG, EMG, etc.) from 197 subjects and is divided into two subsets, SC and ST. SC refers to "Sleep Cohort", which contains data from healthy subjects, and ST refers to "Sleep Troubled", which contains data from healthy subjects. Troubled", which contains PSG data of subjects with sleep disorders. In SC, the PSG of 153 healthy subjects was recorded throughout the night, and in ST, the PSG of 44 subjects with sleep disorders was recorded. In this paper, the data of healthy subjects were chosen. The EEG signals were taken from Fpz-Cz electrode positions, which is a standard electrode layout for measuring EEG activity. The full name is 'Frontopolar (Fpz) - Central (Cz), where Fpz represents the frontopolar region and Cz represents the central region. The sampling rate is 100Hz. According to the literature [6], we can use the EEG database from the Fpz-Cz lead instead of the signals acquired by the C4-A1 and C3-A2 standard electrodes as recommended in the R&K guidelines. The data were segmented into 30-second intervals, and each sleep segment was classified by an experienced sleep physician according to the R&K criteria [3] into stages W, N1, N2, N3, N4, REM, MOVEMENT, and UNKNOWN. Corresponding sleep stage labels were created and recorded in the Hypnogram file for each sleep signal. Due to the lower data volume of the N1 stage, the accuracy of sleep staging for the N1 category is lower compared to other categories. The Hypnogram file contains the annotations of the sleep patterns corresponding to the PSGs. The label "MOVEMENT" stage refers to the subject's body movements, and the "UNKNOWN" stage indicates that the subject's state was not determined [7].

To ensure the consistency of the study data and to exclude the influence of sleep disorders on the study, data from 13 randomly selected individuals in the Fpz-Cz lead of the SC dataset were used for the experimental analyses in this paper. According to the AASM judgment rule [4], N3 and N4 cycles were combined to form N3 phase data, while MOVEMENT and UNKNOWN data were eliminated [7]. The number of each sleep phase obtained is shown in Table 1.

TABLE 1. The amount of data in the sleep stage.

Sleep Period	Wake	N1	N2	N3-N4	REM	Total
Quantity(size)	12434	878	4575	1576	1608	21071
Percentage	59%	4%	22%	7%	8%	100%

B. DATA PROCESSING

1) PREPROCESSING OF EEG SIGNALS

The raw EEG signal is susceptible to external noise during acquisition. To minimize the effect of this noise on the automatic sleep classification model, it is necessary to pre-process the raw EEG signal, as shown in Figure 2.

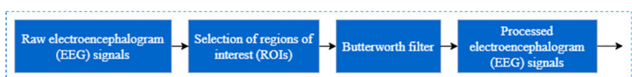

FIGURE 2. EEG pretreatment process.

The Sleep-EDF dataset comprises the polysomnograms of each subject recorded throughout the night and stored on tape using a portable recorder. The polysomnograms of each subject comprised physiological signals from seven channels, as illustrated in Table 2. Since sleep staging was determined using the R&K criteria, which include six classification stages, the markers underwent processing to combine stages S3 and S4. When processing the Sleep-EDF data, it is only necessary to extract the physiological timing signals from the Fpz-Cz channels.

TABLE 2. Polysomnography.

Channel Number	Signal Type	Sampling Frequency	Physical Unit
1	EEGFpz-Cz	100	μV
2	EEGPz-Oz	100	μV
3	EOGhorizontal	100	μV
4	RespOro-nasal	1	—
5	EMGsubmental	1	μV
6	Temprectal	1	DegC
7	Eventmarker	1	—

During sleep, the rhythmic wave frequency range of the generated EEG signals is usually between 0 and 30Hz. To process the raw EEG signals, the filter parameters used in this experiment included a high-pass filter with a cut-off frequency of 0.5Hz and a low-pass filter with a cut-off frequency of 30Hz [8]. The signal was filtered between 0.5Hz and 30Hz using a Butterworth filter with a finite impulse response design [9].

As shown in Figure 3, the original EEG signal maps and the EEG signal maps after filter processing are shown respectively, each for a 10-second time period. By comparing the results of sleep EEG signals before and after preprocessing for the same period, it can be observed that a good denoising effect has been achieved after preprocessing the EEG signals. This provides a reliable basis for subsequent EEG signal analysis and sleep staging.

This preprocessing process helps to remove high and low-frequency noise from the raw sleep signal, thereby improving the accuracy and stability of the subsequent sleep staging model.

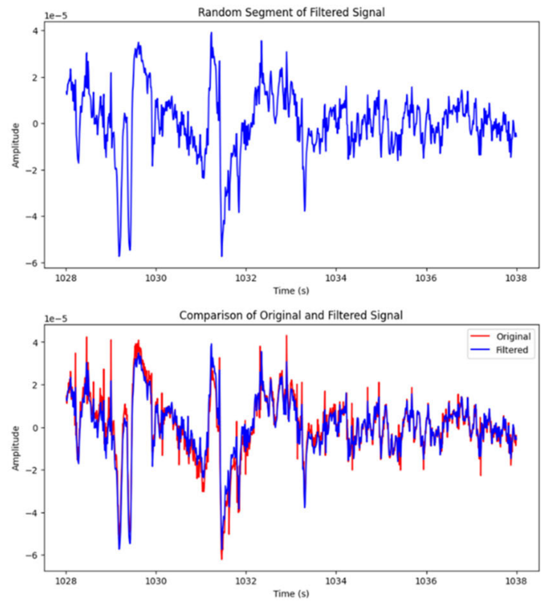

FIGURE 3. Raw EEG signal maps and filtered EEG signal maps.

2) METHODS OF ANALYZING EEG SIGNALS

In the field of EEG, the analysis and comparison of complex sleep EEG signals permits the identification of feature information in a highly recognizable manner. This information is then used to complete the sleep grading process, thus identifying EEG signal features that differ significantly under different sleep stages, and time domain analysis, frequency domain analysis, and time-frequency domain analysis analyze EEG signals more conveniently [10].

Time domain analysis was initially used for the analysis of EEG signals. Quantification of EEG signals is performed through waveform characteristic parameters, including amplitude peak detection, averaging, ANOVA, histogram analysis, and skewness analysis [10]. In addition, for time series signals such as sleep EEG signals, some important features can be directly represented in the time domain, such as the shuttle wave that appears during the N2 period of sleep. However, the information obtainable through time domain methods is extremely limited, as they only consider the resolution of the EEG signal in the time domain and neglect the feature information contained in the frequency domain. Consequently, some researchers have shifted their focus to frequency domain analysis for EEG signal interpretation.

Frequency domain analysis is where the most relevant features of the EEG signal are extracted and the amplitude transform of the signal in the time domain is converted into a power transform in the frequency domain. The commonly used method for this is power spectral estimation, which involves deriving a power spectral estimate from the correlation of a finite dataset with the Fourier transform [11]. However, as the volume of EEG data increases, the effectiveness of power spectral estimation diminishes. Additionally, relying solely on frequency domain analysis results in the loss of time domain resolution, thereby failing to fully characterize the EEG signal.

The frequency information of EEG signals varies in different periods, and neither the use of time domain analysis alone nor the use of frequency domain analysis alone can comprehensively characterize the EEG signals and analyze the information contained in the EEG signals [12]. Therefore, it is necessary to analyze EEG signals by combining the time and frequency domains.

3) VMD AND FFT PREPROCESSING OF EEG SIGNALS

Variational mode decomposition (VMD) is an adaptive and completely non-recursive modal discretization and signal processing method, whose core idea is to construct and solve discretization problems. It determines the number of modal decompositions based on the actual signal, and searches to automatically match the optimal center frequency and finite bandwidth of each mode during the solution process [13]. Thus, achieving successful separation of intrinsic mode functions (IMFs). This function reduces the time series nonstationary and nonlinearity with strong nonlinearity and high complexity.

VMD is a signal decomposition method based on the variational principle, to divide a signal into several linear combinations of local modal functions (IMFs). The basic principle is to find a set of IMFs through an optimization problem such that each IMF is a local frequency mode of the signal and the IMFs are orthogonal to each other. VMD is used for the decomposition of nonlinear and nonsmoothed signals and therefore performs well in dealing with signals containing complex structures.

The Fourier transform is a common method used in signal processing to transform a signal from the time domain to the frequency domain. Discrete Fourier Transform (DFT) is difficult to compute in its discrete form, Fast Fourier Transform (FFT) is an algorithm for grouping and combining sequences based on the period and symmetry of the exponential factors in the DFT, thus reducing the computational space [14]. The FFT does not change the principle of the discrete Fourier Transform but accounts for a significant role in the field of computing. It saves time by reducing the number of multiplicative calculations, especially when the sampling rate is high, and significantly reduces the complexity of the calculations.

The main element of the FFT is that the original sequence of sampling points is unified to form shorter sequences. The symmetry and periodicity of the exponential factors in the DFT formulation allow for the derivation of these shorter sequences and their subsequent combinations, respectively, in a way that reduces the number of multiplications and repeated calculations, thus improving the computational structure [14].

In comparison to the traditional DFT, this algorithm has significant advantages. The DFT has a time complexity of O(N^2) whereas the FFT has been optimized to have a time complexity of O(NlogN), thus making the FFT computationally more efficient in the presence of a large number of sampling points [14]. In addition, FFT became one of the widely used algorithms in computers due to its superiority in computational aspects.

The advantage of spectral density analysis of EEG signals through the FFT function in Python is that the FFT can transform the signal efficiently. FFT is an extremely important method in the field of signal domain processing, frequency domain analysis, and filter variation. Therefore, in this paper, we have chosen to use FFT variations to analyze the EEG signals.

Figure 4. shows that the pre-processing of VMD and FFT is aimed at extracting valid features from the signal data for subsequent sleep staging.

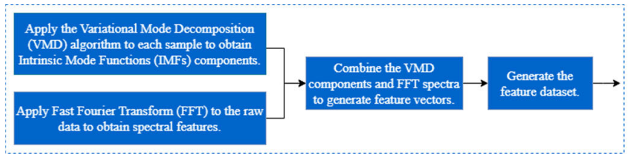

FIGURE 4. VMD and FFT preprocessing flow.

The process starts with inputting the data and then loops through each segment. Within each segment, two processing steps are used: the VMD decomposition and the FFT transform.

1. The original signal is decomposed using the VMD algorithm to obtain IMFs and extract frequency parameters. The basic principle of VMD is to decompose the signal into multiple eigenmode functions corresponding to specific frequency components of the signal. This decomposition effectively captures the local features and frequency domain information of the signal, leading to a more accurate representation of the subsequent features.

2. FFT is performed to obtain its spectral characteristics. The method of converting the signal from the time domain to the frequency domain is to calculate the spectrum of the signal and obtain the energy distribution of the signal at different frequencies. It helps us to understand the frequency components of the signal.

A merging step is performed to superimpose the obtained IMFs and frequency domain features to form the preprocessed feature vector. This step integrates the time-frequency information of the signal, leveraging the strengths of both VMD and FFT to achieve a richer and more comprehensive feature representation. Although the entire feature extraction process is time-consuming, it effectively captures the signal's information and diverse features, thereby aiding the subsequent automatic sleep staging task.

As the process involves frequency domain analysis and mathematical operations, the processing speed is relatively long. Therefore, the effectiveness of the algorithm and the efficiency of its execution need to be considered in its use.

III. MODELLING METHODOLOGY AND ANALYSIS OF RESULTS

A. AUTOMATIC SLEEP STAGING MODEL BASED ON LSTM

In recent years, automatic sleep staging methods based on deep learning have rapidly advanced due to their ability to capture long-term correlations in time-series data. The LSTM model, in particular, utilizes a gating mechanism to adaptively learn and retain information from the input sequence. Hence, its ability to be highly applicable in the analysis of sleep EEG signals. This section aims to explore the LSTM-based automatic sleep staging model and experimentally evaluate its performance [15]. By employing this model, we achieve automatic sleep staging, thereby enhancing the efficiency and accuracy of sleep disorder diagnosis and treatment.

1) LSTM LONG SHORT-TERM MEMORY NETWORK

The Long Short-Term Memory Network is an improved version of the original RNN [16]. It incorporates the concept of cell state, with repeating neural network modules outside the LSTM. The network layer has four layers that regulate and protect the cell state. Figure 5. shows the structure.

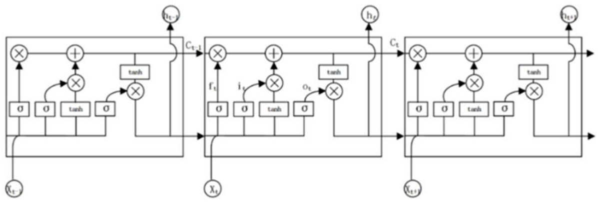

FIGURE 5. LSTM structure.

Each gate is a vector function about the historical state and the current time input, representing the cell state, and is a hidden state vector based on the output. The specific update steps are as follows:

The Forget gate represents the initial stage of the cell update process, in which it is determined whether it is necessary to erase the information in the hidden state before the cell. The gate reads the hidden state of the previous cell and the input of the current cell. Its output is a vector value within 0-1 representing the percentage of information that can be passed through each component.

The Input Gate is the second step, whether it incorporates the updated information into the current unit's state. The input gate receives the hidden state from the previous cell and the current cell's input, producing two components: the input gate vector and the forget gate vector [17]. The input is processed by the Tanh function, and the output is used to update the cell state.

Updating the previous cell state is the third step, in which information related to the previous cell state needs to be evaluated for retention, and data related to the future cell state will be obtained. Combining them is the final cell state at the current moment.

The Output gate is the fourth step that determines the output value of the current cell and runs the control signal generated by the Sigmoid layer to filter the cell states in the output section [17]. The result of processing the cell state through the Tanh function is multiplied by the control signal to obtain the final output value. Eventually, the cell state containing long-term and near-term information will be passed to the next moment.

2) LSTM-BASED AUTOMATIC SLEEP STAGING METHOD

Firstly, there is the input layer, which accepts data of shape (3,3000), where 3 denotes the original dimension plus two extra dimensions, and 3000 denotes a feature dimension of 3000 at each time step. Then there are three bi-directional LSTM layers, each of which has 64 hidden units and is set with "return_sequences=True" so that each time step outputs the complete sequences, which helps to capture long-term dependencies in the time series data. Each bidirectional LSTM layer is followed by a culling layer to prevent overfitting. Finally, a fully connected layer containing five output nodes maps the network output to five categories using a soft-max activation function. Each node corresponds to the probability of one category.

Dropout rate: 0.5. Output layer: 5 nodes, SoftMax activation. Optimizer: Adam, loss: categorical cross-entropy. Model Checkpoint saves the best model based on validation accuracy. Early Stopping stops training if validation accuracy doesn't improve for 10 epochs.

This neural network model is shown in Figure 6.

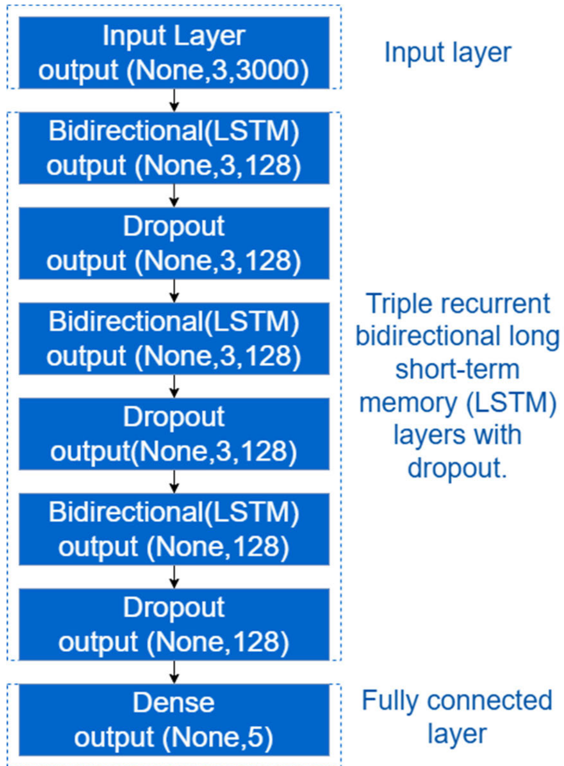

FIGURE 6. LSTM Schematic structure of the automatic sleep staging model.

The detailed parameter design in the model structure is shown in Table 3.

TABLE 3. Detailed parameterization in the model structure.

Layer(type)	OutputShape	Param
bidirectional-al_1(Bidirectional)	(None,3,128)	1569280
dropout_1(Dropout)	(None,3,128)	0
bidirectional_2(Bidirectional)	(None,3,128)	98816
dropout_2(Dropout)	(None,3,128)	0
bidirectional-al_3(Bidirectional)	(None,128)	98816
dropout_3(Dropout)	(None,128)	0
dense_1(Dense)	(None,5)	645

3) ANALYSIS OF EXPERIMENTAL RESULTS

The method had an average accuracy of 0.87.

As shown in Table 4, the results of the experiment show the classification performance metrics for the different sleep stages, including Precision (Pre), Recall (Recall, Re), and F1-Score, as well as the number of samples per category (Support).

1. SleepstageW had the best classification performance achieved with a precision of 0.94, a recall of 0.97, and an F1 score of 0.96. This suggests that the model has high accuracy and recall in predicting wakefulness.

2. Sleepstage1 had the worst classification performance with a precision of 0.24, a recall of 0.12, and an F1 score of 0.16. This may be due to the small sample size or the model's weak classification ability in this category.

3. Sleepstage2 has a better classification performance with a precision of 0.87, a recall of 0.87, and an F1 score of 0.88. The model has a relatively good predictive power for this category.

4. Sleepstage3 and Sleepstage4 had average classification performance, with a precision of 0.80, a recall of 0.65, and an F1 score of 0.72. The model fell slightly short of distinguishing between deep sleep and other stages.

5. SleepstageR had a more average classification performance, with a precision of 0.55, a recall of 0.66, and an F1 score of 0.60. The model had a low classification accuracy in this category and may need further improvement.

TABLE 4. Detailed parameterization in the model structure.

	Precision	Recall	F1-score	Support
SleepstageW	0.94	0.96	0.95	1242
Sleepstage1	0.24	0.12	0.16	66
Sleepstage2	0.87	0.88	0.88	488
Sleepstage3and4	0.80	0.65	0.72	158
SleepstageR	0.55	0.66	0.60	154
accuracy	—	—	0.87	2108
macroavg	0.68	0.65	0.77	2108
weightedavg	0.86	0.87	0.86	2108

The confusion matrix for the LSTM model is shown in Figure 7, presenting the performance of this deep learning model in the experiment. The confusion matrix demonstrates that the model has a predicted value for each category. The values on the diagonal represent the number of correct predictions, while the values on the off-diagonal are the cases of errors. In particular, the model performs more accurately for "SleepstageW" (wakefulness), while for the other categories, "Sleepstage1" has a relatively low recognition accuracy, possibly due to the lack of distinctive features of the category or insufficient data samples.

B. AUTOMATIC SLEEP STAGING MODEL BASED ON LSTM AND CNN

Different machine learning methods have different classification criteria and fitting capabilities, and the accuracy

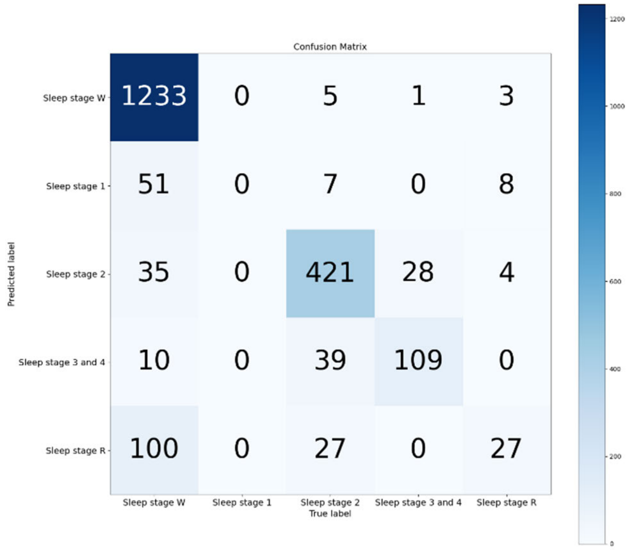

FIGURE 7. Confusion Matrix for LSTM models.

of the classification results of these methods relies on human-selected features. In the field of deep learning, Convolutional Neural Networks (CNNs) can autonomously learn features of the input signal, while Long Short-Term Memory Networks (LSTMs) can selectively learn these features. Combining these two networks could be more effective in exploring the feasibility of automatic sleep staging.

1) AUTOMATIC SLEEP STAGING METHOD BASED ON LSTM AND CNN

This neural network model consists of four key components as shown in Figure 8.

The first is the input layer, which receives data of shape (1+K,3000), where 1+K denotes the number of time steps plus an additional dimension K, and 3000 denotes a feature dimension of 3000 for each time step. This is followed by four convolutional layers, with convolutional kernels of 32, 64, 128, and 256. Each of the convolutional layers is activated using the ReLU activation function, and the padding is set to the "same" to maintain the dimensionality of the input and output.

Then there are two bidirectional LSTM layers, each containing a fixed number of hidden units in the order they appear in the layer_sizes list. Each LSTM layer is set with return_sequences=True, which is used to ensure that the output of each time step is retained, which is significant when working with time series data. By using bidirectional LSTM layers, the model can combine forward and backward information from the time series.

Behind the LSTM layer is a flattened layer that spreads the multidimensionality into one dimension to make the input the fully connected layer. This is followed by two fully connected layers; the first fully connected layer contains 128 neurons and utilizes the ReLU activation function [18]. This is

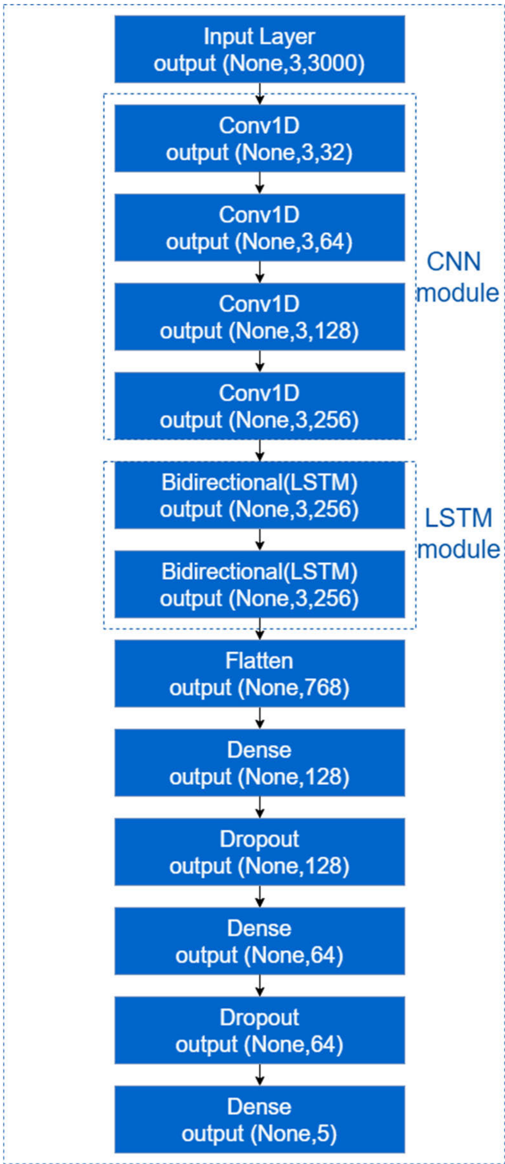

FIGURE 8. Schematic structure of automatic sleep staging model with LSTM and CNN.

followed by a Dropout layer with a 50% packet loss rate set to prevent overfitting. The second fully connected layer contains 64 neurons and again uses the ReLU activation function, followed by another Dropout layer to prevent overfitting [18]. Finally, the output layer is a fully connected layer using a SoftMax activation function. This layer maps the network output to the probabilities of multiple categories, with the output size corresponding to the number of target categories in the model. The design details of the model are shown in Table 5.

The model has 3000 input features and 5 output categories for multi-class classification. It uses two hidden layers with 128 neurons each and a batch size of 100. The Adam optimizer and categorical cross-entropy loss function are used, with accuracy as the evaluation metric.

TABLE 5. Detailed parameterization in the model structure.

Layer(type)	OutputShape	Param
input_1(InputLayer)	(None,3,3000)	0
conv1d(Conv1D)	(None,3,32)	288032
conv1d_1(Conv1D)	(None,3,64)	6208
conv1d_2(Conv1D)	(None,3,128)	24704
conv1d_3(Conv1D)	(None,3,256)	98560
bidirectional(Bidirectional)	(None,3,256)	394240
bidirectional_1(Bidirectional)	(None,3,256)	394240
flatten(Flatten)	(None,768)	0
dense(Dense)	(None,128)	98432
dropout(Dropout)	(None,128)	0
dense_1(Dense)	(None,64)	8256
dropout_1(Dropout)	(None,64)	0
dense_2(Dense)	(None,5)	325

2) ANALYSIS OF EXPERIMENTAL RESULTS

The experimental results achieved a good classification performance with a precision rate of 0.92 and a loss function of 0.33. Table 6 presents the evaluation of classification performance across different sleep stages, detailing the precision, recall, and F1 score, along with the number of samples in each category.

1. SleepstageW had the best classification performance with a precision of 0.97, a recall of 0.98, and an F1 score of 0.98. This suggests that the model showed high precision and recall in predicting wakefulness periods, and was more capable of recognizing this category.

2. Sleepstage1 had the worst classification performance with a precision of 0.59, a recall of 0.36, and an F1 score of 0.45. This may be due to the small sample size or the model's weak ability to classify in this category. It should be noted that a precision and F1 score of 0.00 may indicate that the samples were not classified correctly.

3. Sleepstage2 had a better classification performance, with a precision of 0.91, a recall of 0.90, and an F1 score of 0.88. The model is relatively good at predicting this category, but there is still room for improvement.

4. Sleepstage3 and4 had an average classification performance, with a precision of 0.86, a recall of 0.80, and an F1 score of 0.83. The model was slightly insufficient in distinguishing between deep sleep and the other stages, and more training samples or model tuning was needed.

5. SleepstageR had a relatively average classification performance, with a precision of 0.75, a recall of 0.84, and an F1 score of 0.79. The model's classification accuracy in this category is low, and further improvements may be needed, for example, by adding more features or adjusting the model's parameters to improve performance.

Overall, the weighted average accuracy of the model is 0.92, as shown in Table 6 indicating that the model has a good classification performance overall. However, the macro-averaged (Macroavg) accuracy is 0.81 and the macro-averaged F1 score is 0.78.

As Figure 9 shows the confusion matrix of the LSTM+CNN model, we can deeply analyze the performance

TABLE 6. Detailed parameterization in the model structure.

	Precision	Recall	F1-score	Support
SleepstageW	0.97	0.98	0.98	1242
Sleepstage1	0.59	0.36	0.45	66
Sleepstage2	0.91	0.90	0.9	488
Sleepstage3 and4	0.86	0.80	0.83	158
SleepstageR	0.75	0.84	0.79	154
accuracy	—	—	0.92	2108
macroavg	0.81	0.78	0.79	2108
weightedavg	0.92	0.92	0.92	2108

of the deep learning model in this task. The confusion matrix shows how well the model predicts the five different sleep stages. In particular, the "SleepstageW" category has the most accurate prediction. In contrast, "Sleepstage1" and "SleepstageR" have a lower number of correct predictions.

We can see that the model performs better in dealing with certain categories, while further optimization is needed on other categories. It is suggested that future work could focus on increasing the training samples of difficult-to-identify categories, adjusting the model structure or parameters to reduce overfitting, and exploring new feature extraction techniques to improve the overall performance and robustness of the model. Such improvements will help the model to perform more accurately and reliably in real-world applications.

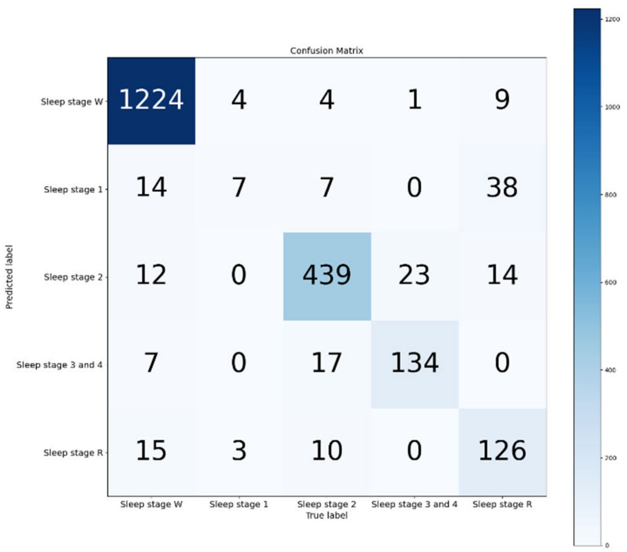

FIGURE 9. Confusion Matrix of the LSTM+CNN model.

IV. EXPERIMENTS AND TEST RESULTS

A. AUTOMATIC SLEEP STAGING MODEL BASED ON LSTM

In this paper, we develop an automatic sleep staging model using two different time-frequency analysis methods: LSTM and LSTM+CNN. To effectively compare the staging performance of these models and determine which provides the best feature characterization, this section will present a comparative analysis of their classification results.

1) INTRODUCTION OF AN EXPERIMENTAL PLATFORM

The platform used for the experiments is Intel(R) Core (TM)i5-10210UCPU@1.60GHz, 2.11 GHz. The platform is configured with a deep learning environment including TensorFlow and Keras. TensorFlow is an open-source software library for building and training machine learning models. It provides tools and APIs for deep learning and machine learning. Keras is integrated into TensorFlow as a high-level API for TensorFlow and can be used within the TensorFlow framework [19]. Keras is designed to be easy to use, making neural network models faster and more intuitive. The preprocessing phase uses Python programs to filter and reduce noise in the EEG data and integrate the datasets. In the classification phase, a CNN program for automatic sleep staging models was written based on Keras. The version of TensorFlow used was 2.9.1 and the version of Keras was 2.6.0.

TensorFlow is our machine learning system built for deep neural networks, which combines optimization techniques in computation to efficiently compute a variety of mathematical expressions that are then used to train and run deep neural networks [19]. These networks can be applied to many scenarios, including image recognition, handwritten digit classification, recursive neural networks, word embedding, natural language processing, video detection, and more.

The Keras neural network framework is a high-level API written in Python that facilitates rapid experimentation and the ability to turn ideas into results quickly. This is an important aspect of conducting effective research.

This code snippet imports a variety of Python libraries and modules essential for data analysis, machine learning, and deep learning. It includes NumPy and Pandas for data manipulation, Matplotlib for visualization, and TensorFlow with Keras for building machine learning models, especially LSTM-based neural networks. Additionally, it leverages Scikit-learn for tasks such as data preprocessing, model evaluation, and computing class weights, while VMD (Variational Mode Decomposition) is used for signal processing. These libraries are often applied in areas like time-series prediction, classification, feature extraction, and tasks in natural language processing, speech recognition, and signal analysis.

The signal data preprocessing using VMD + FFT took approximately 21.32 minutes. The total training time for the CNN + LSTM model was around 20 minutes, while the training time for the LSTM alone was about 33 minutes.

4.1.2. Experimental evaluation indicators

This experiment uses CNNs to test the classification of sleep signal datasets analyzed at different temporal frequencies into the corresponding sleep stages. The dataset is divided into training set and validation and test set in the ratio of 8:1:1 for classification and recognition. Confusion matrix, precision, recall, and F1 value are used to evaluate the classification performance.

In addition to the regular evaluation indicators, two additional evaluation indicators are used:

1. Macro-averagingF1-Score: Assuming that all categories are equally important, the arithmetic mean of the F1 values for each category is calculated and then the macro-averaged F1 values are calculated [20]. Thus, the average performance of the model in each category is presented without considering the effect of category imbalance. This approach is most appropriate when the model's performance in each category is equally important to the task at hand.

2. Weighted-averagingF1-Score: The weighted average F1 score is calculated as a weighted average of the F1 scores for each category, with weights proportional to the true value that each category represents [20]. This implies that the model's performance on larger categories has a greater impact on the weighted average F1 score, as these categories contain a larger number of samples. This approach is well suited to address the problem of category imbalance i.e. some categories may be more important than others.

The formulae for each criterion are as follows:

$$Pre = TP/(TP + FP) \tag{1}$$

$$Re = TP/(TP + FN) \tag{2}$$

$$F1 = (2PR \times RE)/(PR + RE) \tag{3}$$

where TP (True Positive) represents a positive sample with a positive prediction, TN (True Negative) represents a negative sample with a negative prediction, FP (False Positive) represents a negative sample with a positive prediction, and FN (False Negative) represents a positive sample with a negative prediction.

B. COMPARISON OF RESULTS FROM DIFFERENT CLASSIFICATION MODELS

The two classification networks were validated using the Sleep-EDF dataset. Both the LSTM and LSTM+CNN networks used preprocessed signals as input. During the experiments, various network structures were compared to identify the best-performing model.

To visually demonstrate the performance of different networks for sleep staging, this figure compares the classification accuracy (ACC) and loss function (LOSS) of various models, validated against the Sleep-EDF dataset. The performance of the two network models LSTM and LSTM+CNN when processing the same input data is compared in detail in the bar chart.

For the first dataset, the LSTM model achieved an accuracy of 0.87 and a loss value of 0.49. This indicates that although the LSTM model has some advantages in dealing with time-series data, neither its accuracy nor its loss value is optimal.

The second group, the LSTM+CNN model, which combines the time-series processing capability of LSTM with the feature extraction capability of CNN, shows a significant increase in accuracy to 0.92 and a decrease in loss value to 0.33. By combining the strengths of the two networks, the model demonstrates higher accuracy and lower error in the classification task.

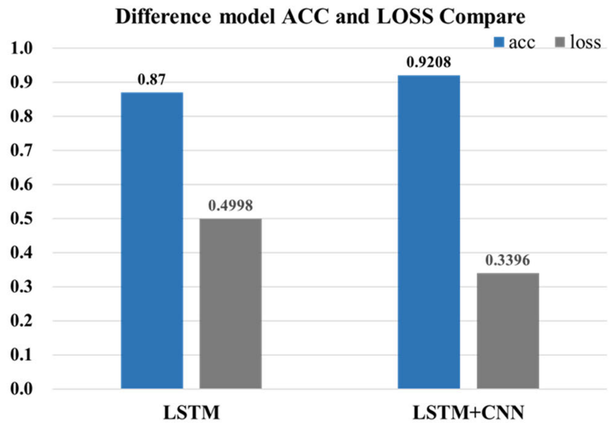

FIGURE 10. Comparison plot of ACC and LOSS for different models.

Figure 11 illustrates a comparison of the performance metrics of the LSTM and LSTM+CNN feature extraction networks, validated using the Sleep-EDF dataset across five sleep stages. The performance of these two networks in each sleep stage was evaluated separately by taking the preprocessed signals as input.

The LSTM+CNN feature extraction network outperforms the use of LSTM models in sleep-stage classification tasks. The LSTM+CNN model can extract complex features from the signals more efficiently and therefore improves the classification accuracy.

This paper shows that the performance of the classification task can be significantly improved by reasonably combining the advantages of different network models. Future research can focus on further optimizing these combined models and exploring additional network structures and data processing methods to enhance the accuracy and robustness of the models.

C. COMPARISON OF RELEVANT RESEARCH PROGRAMMERS AND DISCUSSIONS

To validate the efficacy of the proposed algorithm, comprehensive comparisons are conducted with state-of-the-art sleep staging methodologies published in recent years, as summarized in Table 7. Among these, MMASleepNet [21] (2022) introduced a multimodal attention fusion mechanism using multimodal electrophysiological signals, achieving an accuracy of 89.10%. However, its reliance on diverse physiological inputs limits its practicality in resource-constrained scenarios. In contrast, SleepTransformer [22](2022) adopted a single-channel EEG input and leveraged the self-attention mechanism of the Transformer model, attaining 88.70% accuracy while enhancing interpretability. DeepSleepNet-Lite [23] (2021) prioritized lightweight design and uncertainty estimation for single-channel EEG, though its

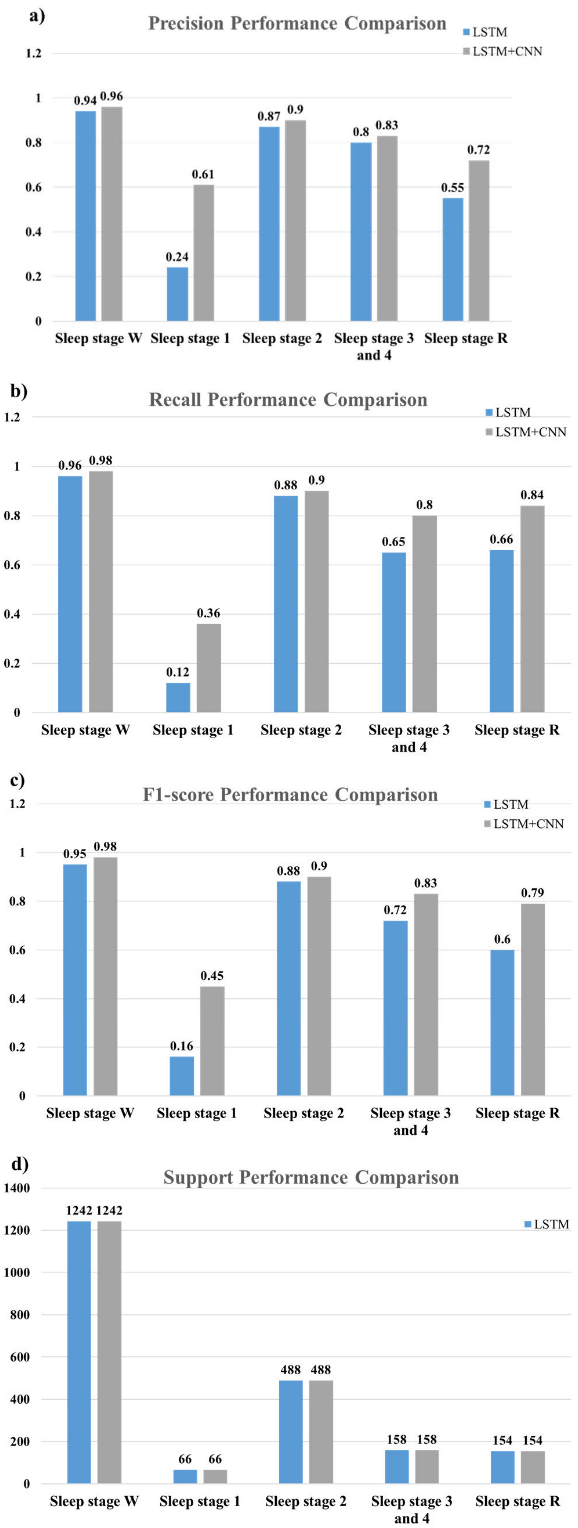

FIGURE 11. Comparison of the average detection accuracy under the feature extraction network for both signals (a) Model precision performance comparison; (b) Model recall performance comparison; (c) Model f1-score performance comparison; (d) Model support performance comparison.

accuracy remained relatively low at 85.30%. SleepVST [24] (2024) achieved cross-modal transfer learning by integrating video and physiological signals, yielding a higher accuracy of 91.20%, yet its dependency on video data complicates deployment. EdgeSleepNet [25] (2024) optimized single-channel EEG staging through dynamic pruning and edge computing, reaching 89.30% accuracy but with potential trade-offs in feature granularity.

TABLE 7. Comparison of the accuracy of the methods in this paper with the accuracy of sleep staging methods in the last three years.

Methodologies	Time Limit	Input Type	ACC (%)	Innovation point
MMASleep Net[21]	2022	Multimodal physiological signals	89.10	Multimodal attention fusion
Sleep Transformer[22]	2022	Single channel EEG	88.70	Self-attentional interpretability
DeepSleep Net-Lite**Error! Reference source not found.**	2021	Single channel EEG	85.30	Lightweight and uncertainty estimation
SleepVST[24]	2024	Video and physiological signals	91.20	Cross-modal transfer learning
EdgeSleepNet[25]	2024	Single channel EEG	89.30	Dynamic pruning and edge optimization
This article	—	Single channel EEG	92.07	VMD and FFT preprocessing and LSTM+CNN hybrid architecture

The proposed method distinguishes itself through two key innovations. First, it incorporates VMD-FFT joint preprocessing, which effectively decomposes and denoises single-channel EEG signals while preserving critical spectral features, addressing noise sensitivity issues prevalent in prior works. Second, a hybrid LSTM+CNN architecture is designed to synergistically capture long-term temporal dependencies via LSTM and local spatial patterns through CNN, outperforming both pure attention-based [22] and lightweight CNN frameworks [23], [25]. With an accuracy of 92.07% using only single-channel EEG inputs, this approach surpasses all recent benchmarks, including multi-modal [21], [24] and computationally intensive models [22], demonstrating superior generalizability and efficiency. The integration of signal decomposition, spectral enhancement, and dual-domain feature learning establishes a new state-of-the-art for practical, high-precision sleep staging systems.

Despite its promising performance, this study has two notable limitations. First, while our framework achieves high accuracy, its statistical validation could be further strengthened through advanced resampling methods and hypothesis testing, which were not fully explored due to dataset size constraints. Second, the absence of cross-validation analyses, such as leave-one-subject-out validation, limits our ability to conclusively assess generalizability across diverse demographic and clinical subgroups. Future efforts will prioritize these aspects to ensure rigorous validation and clinical translatability.

V. EXPERIMENTS AND TEST RESULTS

This paper introduces a novel automatic sleep staging framework that integrates VMD and FFT preprocessing with a hybrid LSTM+CNN architecture, achieving state-of-the-art accuracy of 92.07% using single-channel EEG signals. Our method addresses key limitations in existing approaches through two core innovations. First, the VMD and FFT joint preprocessing technique decomposes raw EEG signals into intrinsic mode functions via VMD, followed by spectral enhancement using FFT. This dual-step process effectively mitigates noise interference while preserving critical frequency-domain features, outperforming conventional spectral methods used in prior works. Second, the LSTM+CNN hybrid model uniquely combines the temporal modeling capability of LSTM networks with the spatial feature extraction strength of CNNs. This architecture captures both long-range sleep stage transitions and localized EEG patterns, surpassing pure attention-based Transformers and lightweight CNNs in accuracy.

Although the model performs well in some classification categories, in specific categories, such as Sleep stage 1, the classification is poor and needs to be improved.

Research for the future can be carried out in the following aspects: further optimizing the model structure and algorithms to improve the accuracy and generalization ability of the automatic sleep staging model; exploring more advanced EEG signal processing methods to enrich the feature expression and improve all the performances of the model; and combining the multimodal data, including ECG, respiration and so on, to further improve the validity of sleep staging. When conditions permit, the model is used in clinical practice to judge whether it is feasible. Future studies can enhance the automatic sleep staging model to offer greater support and assistance in the diagnosis and treatment of sleep disorders. Applying Grad-CAM in analyzing model behavior and interpreting predictions. We wholeheartedly agree with the reviewer's concerns and acknowledge their critical role in advancing sleep staging research. These limitations will guide our subsequent investigations, where we will rigorously implement statistical validation frameworks and cross-validation strategies. We further pledge to publicly share code and collaborate with clinical partners to acquire larger, multi-center datasets, thereby addressing these methodological gaps while maintaining transparency and reproducibility.

REFERENCES

[1] K. Harding and M. Feldman, "Sleep disorders and sleep deprivation: An unmet public health problem," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 47, no. 4, pp. 473–474, 2008.

[2] K. R. Reddy, P. N. H. Vardhan, S. Pary, S. Sharma, and M. Sanjay, "A CNN-LSTM model for sleep stage scoring using EEG signals," in *Proc. Int. Conf. Commun., Circuits, Syst. (ICS)*, Bhubaneswar, India, May 2023, pp. 1–6, doi: 10.1109/IC3S57698.2023.10169177.

[3] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Los Angeles, CA, USA: Brain Information Service, Brain Research Institute, Univ. California, 1968.

[4] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan, S. Redline, K. P. Strohl, S. L. D. Ward, and M. M. Tangredi, "Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events," *J. Clin. Sleep Med.*, vol. 8, no. 5, pp. 597–619, Oct. 2012.

[5] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, Jun. 2000.

[6] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1185–1194, Sep. 2000.

[7] J. Chen, L. Zhou, C. Jiang, Z. Chen, L. Zhang, H. Zhou, W. Kang, X. Jiang, Y. Li, N. Luo, M. Yao, M. Niu, S. Chen, X.-N. Zuo, L. Li, and J. Liu, "Impaired ocular tracking and cortical atrophy in idiopathic rapid eye movement sleep behavior disorder," *Movement Disorders*, vol. 37, no. 5, pp. 972–982, May 2022.

[8] Z. Yue, "Research on the preprocessing of multilead EEG signal and its sleep staging based on multilead EEG signal," Ph.D. dissertation, School Bioinformation, Chongqing Univ. Posts Telecommun., Chongqing, China, 2017.

[9] D. Kubanek, J. Koton, J. Jerabek, and D. Andriukaitis, "(N + α)-order low-pass and high-pass filter transfer functions for non-cascade implementations approximating Butterworth response," *Fractional Calculus Appl. Anal.*, vol. 24, no. 3, pp. 689–714, Jun. 2021.

[10] L. Wei, "Research on automatic sleep staging method combining time-frequency information and deep learning," Ph.D. dissertation, School Comput. Inf. Technol., Beijing Jiaotong Univ., Beijing, China, 2018.

[11] R. Aldemir, E. Demirci, H. Per, M. Canpolat, S. Özmen, and M. Tokmakçi, "Investigation of attention deficit hyperactivity disorder (ADHD) subtypes in children via EEG frequency domain analysis," *Int. J. Neurosci.*, vol. 128, no. 4, pp. 349–360, Apr. 2018.

[12] H. Xu, Z. Pei, Q. Han, M. Hou, X. Qian, T. Weng, Y. Tian, Z. Qiu, and B. Zhou, "MASTF-net: An EEG emotion recognition network based on multi-source domain adaptive method based on spatio-temporal image and frequency domain information," *IEEE Access*, vol. 12, pp. 8485–8501, 2024.

[13] X. Qin, D. Xu, X. Dong, X. Cui, and S. Zhang, "EEG signal classification based on improved variational mode decomposition and deep forest," *Biomed. Signal Process. Control*, vol. 83, May 2023, Art. no. 104644.

[14] S. Zhang, "Application of FFT algorithm in the analysis of the spectral density of brainwave signals in the implementation," *Comput. Knowl. Technol.*, vol. 13, no. 13, pp. 225–227, 2017.

[15] L. Zhuang, M. Dai, Y. Zhou, and L. Sun, "Intelligent automatic sleep staging model based on CNN and LSTM," *Frontiers Public Health*, vol. 10, Jul. 2022, Art. no. 946833.

[16] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.

[17] R. Gao, Y. Tang, K. Xu, Y. Huo, S. Bao, S. L. Antic, E. S. Epstein, S. Deppen, A. B. Paulson, K. L. Sandler, P. P. Massion, and B. A. Landman, "Time-distanced gates in long short-term memory networks," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101785.

[18] Y. U. Khan, A. R. Hassan, and S. N. Yu, "EEG-TCNet: An end-to-end temporal convolutional network for mobile EEG-based sleep staging," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 2, pp. 722–732, Feb. 2023, doi: 10.1109/TBME.2022.3201238.

[19] J. Wang, A. B. Johnson, C. D. Martinez, E. F. Chen, G. H. Kim, I. J. Smith, and K. L. Brown, "Diagnostic performance of artificial intelligence-assisted PET imaging for Parkinson's disease: A systematic review and meta-analysis," *NPJ Digit. Med.*, vol. 7, no. 1, pp. 1–11, Jan. 2024, doi: 10.1038/s41746-024-01012-z.

[20] M. A. Sadeghi, D. Stevens, S. Kundu, and A. Smith, "Detecting Alzheimer's disease stages and frontotemporal dementia in time courses of resting-state fMRI data using a machine learning approach," *J. Imag. Informat. Med.*, vol. 37, pp. 1–16, 2024.

[21] Z. Yubo, L. Yingying, Z. Bing, Z. Lin, and L. Lei, "MMASleepNet: A multimodal attention network based on electrophysiological signals for automatic sleep staging," *Frontiers Neurosci.*, vol. 16, Aug. 2022, Art. no. 973761.

[22] H. Phan, K. Mikkelsen, O. Y. Chén, P. Koch, A. Mertins, and M. De Vos, ''SleepTransformer: Automatic sleep staging with interpretability and uncertainty quantification,'' *IEEE Trans. Biomed. Eng.*, vol. 69, no. 8, pp. 2456–2467, Aug. 2022.

[23] L. Fiorillo, P. Favaro, and F. D. Faraci, ''DeepSleepNet-lite: A simplified automatic sleep stage scoring model with uncertainty estimates,'' *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 2076–2085, 2021.

[24] J. F. Carter, J. Jorge, O. Gibson, and L. Tarassenko, ''SleepVST: Sleep staging from near-infrared video signals using pre-trained transformers,'' 2024, *arXiv:2404.03831*.

[25] J. Shin, S. Gwak, S. J. Shin, and S. Bang, ''Simultaneous estimation and variable selection for a non-crossing multiple quantile regression using deep neural networks,'' *Statist. Comput.*, vol. 34, no. 3, pp. 1–15, Jun. 2024, doi: 10.1007/s11222-024-10418-4.

LIYING ZHU received the Bachelor of Engineering degree in biomedical engineering from Guangdong Medical University, Dongguan, Guangdong, China, in 2024.

She was a Research Intern with the National Innovation Center for Advanced Medical Devices, from 2023 to 2024. She has a keen interests in biomedical signal processing and control, with a focus on contributing to interdisciplinary research in signal and image measurement and analysis within clinical medicine and biological sciences. Her work centers on advancing practical applications that utilize methods and devices for clinical diagnosis, patient monitoring, and management.

QINGXIN GUAN received the Bachelor of Engineering degree in materials science and engineering from Zhuhai College, Jilin University, Zhuhai, China, in 2024. Her primary field of study is medical materials.

She was an Assistant Researcher with the Supramolecular Materials Research Institute, Jilin University, Zhuhai, from 2023 to 2024. During her internship, she gained valuable experience in the analysis of EEG signals, which contributed to her understanding of the intersection between materials science and medical applications. Her current research interest includes the development and characterization of advanced biomedical materials.

YUNQING LIU received the bachelor's degree in intelligent medical engineering from Guangdong Medical University, in 2024.

Her current research interests include the moving cascaded acoustic holographic algorithm based on physics-driven networks and the applications of artificial intelligence in the optimization method of acoustic holography.

JIAYUE XU received the B.S. degree in intelligent medical engineering from Guangdong Medical University, Guangdong, China, in 2024. Her current research interests include material microstructural detection and analysis based on optical coherence tomography technology and multi-diffraction order wide-band inorganic compound detection.

SHA MA received the bachelor's and Ph.D. degrees in digital communications from the University of Central Lancashire, U.K., in 2006 and 2011, respectively.

She is currently a Lecturer with the School of Biomedical Engineering, Guangdong Medical University. She is a member of the Board of Directors of the Intelligent Rehabilitation Technology and Transformation Branch, Guangdong Rehabilitation Medical Association. Her research interests include digital signal and image processing, virtual reality, and rehabilitation system development.