MGSleepNet: A Multi-Granularity Sleep Staging Network Based on EEG and EOG Signals

Zhentao Huang[1], Yuhao Jiang[2], Yin Tian[1,2,*]
1. School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
2. School of Life and Health Information Science and Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
* Corresponding author: tianyin@cqupt.edu.cn

Abstract

Sleep staging is the foundation of sleep analysis. Recent studies have attempted to integrate multimodal signals, such as electroencephalogram (EEG) and electrooculogram (EOG), to enhance the sensitivity of models. However, these attempts still face limitations in effectively merging multimodal signals, particularly in capturing the interplay of global and local information during sleep stages simultaneously. To address this issue, we propose a Multi-granularity Sleep Staging Network (MGSleepNet), which integrates two core modules: the Global Feature Integration module (GFI) and the Fine-grained Information Capture module (FIC). The GFI effectively captures the global features of EEG and EOG signals through multi-scale convolution, channel attention mechanisms, and spatial attention mechanisms. The FIC module obtains fine-grained interaction information between EEG and EOG by segmenting time periods and employing cross-attention mechanisms. The combination of these modules resulted in an accuracy of 83.16% and 82.46% in five-fold cross-validation across subjects on the Sleep-edf-20 and Sleep-edf-78 datasets, respectively. In addition, we also achieved better than opportunity level performance on our own data sets. Finally, the ablation studies confirmed the benefits of integrating global and fine-grained relevance paradigms to enhance sleep staging performance. The model input research indicated that MGSleepNet demonstrates good performance in sleep staging outcomes.

Keywords: Sleep signal；EEG；EOG；Bi-LSTM；

Introduction

Sleep accounts for about one-third of a person's life and plays a crucial role in maintaining overall human health. The quality of sleep directly affects physical and mental well-being, and sleep-related disorders have become a major health issue worldwide. Therefore, monitoring sleep, analyzing sleep structure, and assessing sleep quality are considered primary areas of research interest. Among these, sleep staging is a key aspect of evaluating sleep structure, making it particularly important for diagnosing and treating various sleep disorders. Through accurate sleep staging, healthcare professionals can better understand patients' sleep conditions, thereby developing more effective intervention measures.

Polysomnography (PSG) is the gold standard for clinically diagnosing sleep disorders, recording multiple physiological signals such as electroencephalogram (EEG), electrocardiogram (ECG), electromyogram (EMG), electrooculogram (EOG), blood oxygen saturation, and respiration. PSG recordings are typically divided into 30-second segments, with each segment's data analyzed by experts according to specific rules to classify the sleep stages. In particular, PSG or single-channel EEG recordings are usually segmented into 30 seconds, with each segment manually examined by sleep specialists and then categorized into five types according to the American Academy of Sleep Medicine (AASM) rules: awake (W), rapid eye movement (REM), and non-rapid eye movement (N1, N2, N3). However, the manual staging of PSG by experts in the field is both time-consuming and labor-intensive, heavily reliant on the physician's expertise, and prone to subjective bias.

Deep learning has made significant progress in various fields such as image processing and natural language processing. Consequently, research on automatic sleep staging using deep learning methods has gained immense popularity and made substantial advancements. Despite the growing interest in deep learning-based approaches, the performance of automatic sleep staging methods based on limited data is currently unsatisfactory. Firstly, there is the challenge of extracting effective features from raw electroencephalogram (EEG) signals. According to the AASM sleep standards, the physiological signals of different sleep stages typically exhibit different characteristic waveforms, and EEG is time-dependent; thus, many researchers have adopted spatiotemporal combination models to extract data features. For instance, Bao et al.(Bao et al., 2024) proposed the Feature Fusion Temporal Convolutional Network (FFTCN) for automatic sleep staging using single-channel EEG data. This algorithm employs a one-dimensional convolutional neural network (1D-CNN) to extract temporal features from the raw EEG and utilizes a two-dimensional CNN (2D-CNN) to extract time-frequency features from the spectrogram generated at the temporal level through continuous wavelet transform (CWT). Subsequently, these features are integrated and further input into a Temporal Convolutional Network (TCN) to classify sleep stages at the sequence level. Lee et al. (Lee et al., 2024) proposed a deep learning structure composed of multi-kernel convolutional neural networks and bidirectional long short-term memory for sleep stage classification. Li et al. (Li et al., 2024) introduced a new method called SleepFC. This method includes a Convolutional Feature Pyramid Network (CFPN), Cross-Scale Temporal Context Learning (CSTCL), and a classification network based on a Class Adaptive Fine-Tuning Loss Function (CAFTLF). Furthermore, there is a failure to adequately capture the cross-modal contextual relationships between different modalities. In sleep staging, electroencephalography (EEG) is the most commonly used signal; however, EOG can detect eye movements, which are fundamental indicators for distinguishing between Rapid Eye Movement (REM) and Slow Eye Movement (SEM) (Pradeepkumar et al., 2024). Additionally, the EEG waves during REM and N1 stages are similar, while the EOG waves differ significantly. On the other hand, the classification of N2 and N3 stages primarily relies on the prominent waves in the EEG. However, most current studies focus solely on EEG signals (An et al., 2021;McCloskey et al., 2019), neglecting relevant information from other modal data, making it challenging to explore comprehensive task knowledge. Although some research efforts utilize information obtained from multiple modalities for sleep staging (Lin et al., 2023;Jia et al., 2022), unfortunately, they do not adequately consider the uniqueness of each modality and the differences between them. Furthermore, current methods based on electroencephalography do not sufficiently take into account the characteristic waves and features defined in sleep staging criteria, failing to extract the fine-grained frequency characteristics specific to each sleep stage. For instance, sleep spindles typically last from 0.5 to 3 seconds with a frequency range of 11 to 16 Hz, while K-complexes last from 0.5 to 1.5 seconds with a frequency of less than 1 Hz (Yücelbaş et al., 2024). These waves are usually transient and momentary, lasting only a few seconds to tens of seconds. However, the length of sleep periods is typically 30 seconds or longer, making it difficult to fully capture their details over such an extended time span.

To address the aforementioned issues, this paper proposes a new Multi-granularity Sleep Staging Network (MGSleepNet) that utilizes EEG and EOG signals for sleep staging, aimed at the automatic classification of sleep stages from multimodal signals. First, we employ a Global Feature Integration (GFI) module, which adopts a deep architecture based on Convolutional Neural Networks (CNN) to thoroughly explore the core features within EEG and EOG signals. By utilizing a multi-scale convolutional architecture, we extract both high and low-quality features from the signals. Additionally, by applying channel and spatial attention modules, our method can identify complex correlation patterns between EEG and EOG signals throughout the entire recording period. Secondly, we designed a Fine-grained Information Capture (FIC) module, which leverages a cross-attention mechanism to capture subtle and local correlations between segments of EEG and EOG signals. "This not only helps in understanding the signal characteristics at a specific point in time, but also combines information from previous time periods to provide a richer and more accurate analytical foundation for the current moment. Furthermore, the results obtained from these two modules are combined and input into the Bidirectional Long Short Term Memory Network (Bi-LSTM) time series module to extract time series features, ultimately resulting in the classification outcomes after passing through a fully connected layer."

Based on the above analysis, we propose a multi-interactive attention model named MGSleepNet, which utilizes EEG and EOG signals for sleep staging. The main contributions of this paper are as follows:

1) We have built a GFI module that first uses a CNN-based multi-scale convolutional network to learn the frequency characteristics of EEG and EOG signals, and then employs channel and spatial attention mechanism modules to discover the global correlation patterns of EEG and EOG across epochs.

2) We designed an FIC module that utilizes cross-attention to discover fine-grained correlations between subtle segments of EEG and EOG within an epoch. While analyzing the current signal segment, it also integrates relevant information from previous signal segments to enhance the understanding and extraction of temporal features of the signals.

3) We conducted experiments on the widely used Sleep-edf dataset to validate the effectiveness and feasibility of our method. Compared to the baseline models, our model achieved better performance, and the ablation studies confirmed the necessity and effectiveness of the modules in MGSleepNet. Finally, we also validated the model on a dataset we built in the lab, yielding favorable results.

Methods

Architecture Overview

Figure 1 reveals an overview of the MGSleepNet model proposed in this study. The model consists of three core modules, namely: 1) Global Feature Integration module, 2) Fine-grained Information Capture module, and 3) Bidirectional Long Short-Term Memory. First, the GFI module is responsible for extracting global features from the EEG and EOG signals within 30-second data segments using multi-scale convolution, channel attention mechanisms, and spatial attention mechanisms. Second, the FIC module captures fine-grained features by segmenting the 30-second data into smaller pieces, specifically extracting interaction information between EEG and EOG segments using cross-attention mechanisms. Third, this study introduces the BI-LSTM module to extract temporal features from the signals. Ultimately, the results of sleep stage classification are output through two fully connected layers. In the subsequent sections, the functions and implementations of each module will be elaborated in detail.

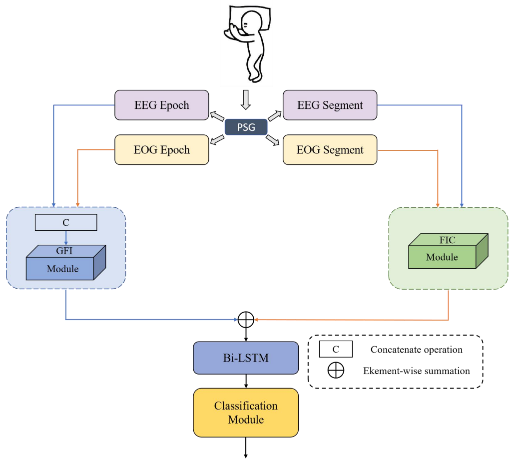

Figure 1: MGSleepNet model framework.

Global Feature Integration module

We have built the Global Feature Integration module (GFI) submodule to extract global interaction information. As shown in Figure 2, GFI mainly consists of three modules: multi-scale convolution module, channel attention module, and spatial attention module. EEG epochs and EOG epochs processed through the GFI module can extract the temporal-frequency, channel, and spatial features inside. The multi-scale CNN is composed of both large-scale CNN and small-scale CNN, which extract low-frequency features and high-frequency features from the signals, respectively. In Figure 2, Conv1d(64, 50, 20) refers to a one-dimensional convolution layer with 64 filters, a kernel size of (1, 50), and a stride of 20. Similarly, Maxpooling(4,4) refers to a max pooling layer with a kernel size of (4, 4). Given the feature map $F \in R^{C \times D}$, where C and D represent the channel dimension and feature dimension, respectively. The channel attention block and spatial attention block are illustrated in formulas (1) and (2), which sequentially extract global channel interaction information *Fc* and spatial interaction information *Fs*. They can be represented as where *Mc* and Ms denote the global channel and spatial attention maps, respectively, and ⊗ signifies element-wise multiplication.

$$F_c = M_c(F) \otimes F \tag{1}$$

$$F_{c,s} = M_s(F_c) \otimes F_c \tag{2}$$

For the channel attention block, this paper employs the SE (Squeeze-and-Excitation) attention mechanism module, which is a lightweight model designed to enhance the feature representation capability of convolutional neural networks by adaptively re-weighting channel feature maps. The process begins with global average pooling (the Squeeze step), which compresses the spatial dimensions of each input feature map to extract a single value that represents the global information of that feature map. Subsequently, in the Excitation step, a small sub-network composed of two fully connected layers processes this global information, involving dimensionality reduction, ReLU activation, dimensionality expansion, and Sigmoid activation, generating weight coefficients corresponding to each channel to evaluate the importance of each feature map. Finally, in the adjustment step, the original feature maps are multiplied by the computed weight coefficients, thereby amplifying the influence of important feature maps while diminishing the effects of less important features.

The spatial attention mechanism model aims to enhance the focus on specific spatial locations in feature maps within convolutional neural networks. First, the model compresses the input channel feature map into a single channel through a convolutional layer, which serves to extract key spatial information from the input feature map. Next, the model applies a Sigmoid activation function to convert the output of the convolutional layer into values between 0 and 1, generating a weight map that represents the importance of each spatial location. In this weight map, values close to 1 indicate that the features at corresponding locations are very important, while values close to 0 suggest that these features are relatively unimportant. Finally, the original input feature map is multiplied by the generated attention weight map, thereby re-weighting the original feature map: enhancing the information in important areas while suppressing unimportant regions.

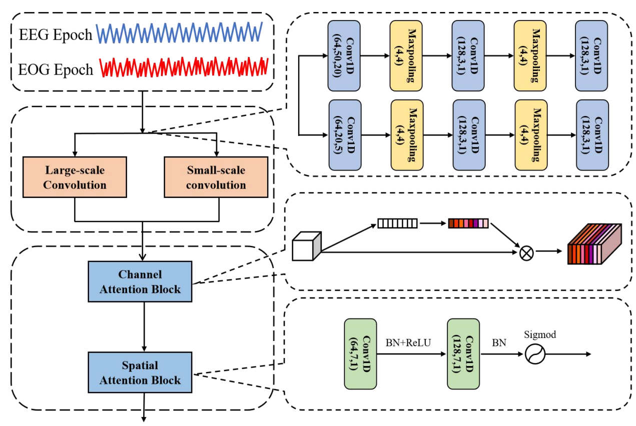

Figure 2: Global Feature Integration module.

Fine-grained Information Capture module

In order to extract instantaneous features of sleep signals, we designed the Fine-grained Information Capture (FIC) module to extract the fine-grained interaction information F_k between segments of EEG and EOG in each epoch pair. As shown in Figure 3, FIC first divides an epoch of the sleep cycle into K equal segments along the time dimension using a fixed time window, represented as $s_i = [p_i^1 \dots p_i^K]$. For a segment $p_i^k = (p_i^{k,1}, p_i^{k,2})$, where $p_i^{k,1}$ and $p_i^{k,2}$ represent the current EEG and EOG signals, respectively, the FIC module first creates new EEG and EOG vectors $p_i^{k'} = (p_i^{k'1}, p_i^{k'2})$ by integrating the previous interaction information F_{K-1}, with the preceding EEG segment and EOG vector represented as p_i^{k-1}. This process enhances the influence of historical information when extracting features from the current signal segment.

Next, $p_i^{k'}$ generates a feature map $\tilde{p}_i^{k'}$ after passing through a 1D convolutional layer. Finally, $\tilde{p}_i^{k'}$ is input into the *k-th* cross-attention block to generate the interaction information F_k. This process can be described by formula (3-4), where Att(.) denotes the cross-attention operation.

$$p_i^{k'} = p_i^k \oplus F_{k-1} \tag{3}$$

$$F_k = Atten(\tilde{p}_i^{k'}) \tag{4}$$

In the FIC module, the k-th cross attention block takes vector $p_i^{k'} = (p_i^{k'1}, p_i^{k'2})$ as input. The Query and Key are assigned values of $\tilde{p}_i^{k',1}$ and its transpose respectively, while the Value is assigned a value of $\tilde{p}_i^{k',2}$. The pairwise similarity between Query and Key is computed using the dot product scaled by a factor of $1/\sqrt{d}$, as shown in equation (5). Subsequently, these obtained weights are distributed to the Value through element-wise multiplication.

$$\delta = soft\max(\frac{\tilde{p}_i^{k',1} \times \tilde{p}_i^{k',2}}{\sqrt{d}}) \tag{5}$$

The FIC module not only integrates multimodal information but also learns the interaction information between EEG and EOG segments, as well as some time series information.

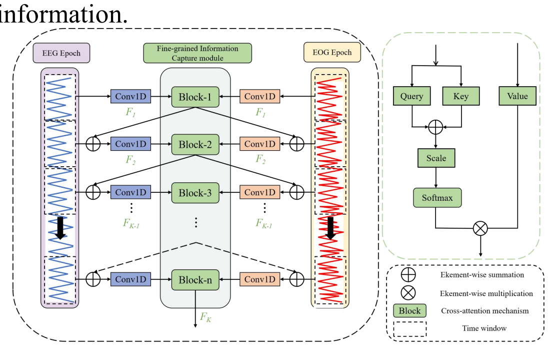

Figure 3: Fine-grained Information Capture module.

Bidirectional long short term memory

Although Convolutional Neural Networks (CNNs) excel in feature extraction, they lack the ability to handle the memory of time series patterns. Therefore, we added a Bidirectional Long Short Term Memory (Bi-LSTM) network at the end of the model, which can maintain short-term memory while also capturing long-term memory states, a crucial aspect for understanding time series data such as EEG and EOG. The Bi-LSTM combines two LSTM layers and can analyze the sequence in both forward and backward directions, allowing the model to understand not only the current and past data patterns but also consider future information segments for a more comprehensive understanding of the entire sequence. The Bi-LSTM network is illustrated in Figure 4, where the Bi-LSTM calculates the entire output ht based on equation (6).

$$h_t = \sigma(W_h \times [\overrightarrow{h}_t, \overleftarrow{h}_t] + b_h) \tag{6}$$

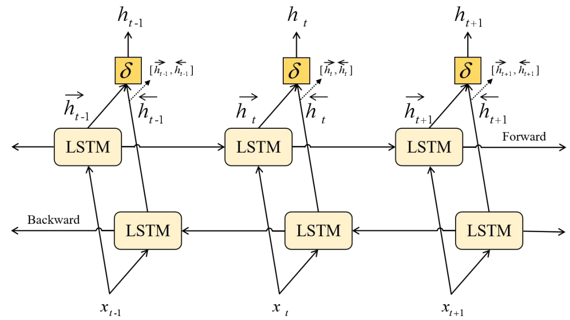

Figure 4: Bidirectional long short term memory module.

Experiment And Results

Datasets

1) Sleep-edf-20(Kemp et al., 2000): The Sleep-edf-20 database is a subset of the Sleep-edf-78 database, representing the 2013 version of the sleep-edf database. The Sleep-edf-20 database now includes overnight polysomnography from 61 healthy subjects and those with mild sleep onset difficulties, along with expert annotations on sleep stages. The recordings are from white males and females aged 21-35 years, who have not taken any medications; they include horizontal EOG, Fpz-Cz, and Pz-Oz EEG, each sampled at a frequency of 100 Hz. In our study, we utilized both dual-channel EEG signals and single-channel EEG signals. A more detailed introduction regarding the interpretation and handling of sleep stage labels is provided in the context of the Sleep-edf-78 database.

2) Sleep-edf-78(Ary et al., 2000): The Sleep-edf-78 database consists of 197 all-night polysomnography records, including EEG, EOG, chin EMG, and event markers. The polysomnographic recordings are labeled with sleep stages W, R, 1, 2, 3, 4, M (movement time), and unknown stages. All polysomnographic recordings were manually scored by trained technicians according to the Rechtschaffen and Kales manual from 1968. We converted this into the sleep stage scoring rules defined by AASM by merging S3 and S4 into a single N3 stage, removing data labeled with Unknown stage tags, and retaining only the data from the 30 minutes before sleep onset and the 30 minutes after sleep offset as usable data. In these two studies, the sampling frequency for EOG and EEG signals was 100 Hz, while the sampling frequency for event marker signals was 1 Hz.

3) Laboratory-built Dataset: To support our research project and acquire high-quality data, we established a non-public database in our laboratory in 2024. This self-built dataset is specifically designed for sleep research and contains six clinical polysomnography (PSG) recordings collected from university students during nighttime sleep. Each PSG recording was conducted in a strictly controlled experimental environment. We selected six healthy university students as research subjects, ensuring that their lifestyle habits and daily routines remained relatively stable during the data collection period to minimize external

factors affecting sleep quality. Each participant underwent a thorough health examination and signed an informed consent form, ensuring that they fully understood the purpose and process of the experiment. The PSG recordings were annotated by qualified sleep specialists according to AASM standards. Specifically, the PSG data for each participant includes electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and other physiological signals (such as electrocardiogram (ECG), respiratory flow, etc.). These multimodal data provide comprehensive physiological information, aiding in the in-depth analysis of sleep structure and quality.

Experiment Setup

In this study, to accurately assess the model's performance, we employed a five-fold cross-subject cross-validation method for a comprehensive analysis of two publicly available datasets. Specifically, we randomly divided the participants in each dataset into five groups, ensuring that each group reflected the characteristics of the overall sample, thereby maximizing data utilization efficiency and accurately evaluating model performance. This approach not only enhanced the reliability of the results but also effectively reduced the impact of individual differences on the model assessment. During model training, we selected the Adam optimizer with a learning rate of 0.001, and to address the class imbalance present in the dataset, we used a weighted cross-entropy loss function to ensure that each class of samples received appropriate attention. The batch size set during training was 128, and the entire training process lasted for 100 epochs. For the evaluation of the model, we comprehensively utilized five evaluation metrics: accuracy, mean F1 score, Cohen's Kappa coefficient, sensitivity, and precision, to thoroughly measure the model's performance.

Comparison Method

During the development and evaluation of the MGSleepNet model, we focused on comparative analysis with a range of current state-of-the-art methods and established baseline models. Given that our dataset encompasses characteristics of both public and private datasets, we utilized only the public datasets for the comparative analysis to ensure transparency and reproducibility, while leveraging the results from the private datasets to further validate the performance of our model. The model proposed in this study was compared with the following baseline methods:

U-time(Perslev et al., 2019) is a time-based fully convolutional network built on the U-Net architecture, specifically designed to map input sequences of arbitrary length to category label sequences on a user-customizable time scale.

ResnetLSTM(Sun et al., 2018) is a deep convolutional network model used for automatic sleep stage classification based on neurophysiological signals. This model increases the network's depth by incorporating residual modules to extract multi-level features of sleep stages, while Long Short-Term Memory (LSTM) networks are employed to learn the sleep transition mechanisms during the sleep process.

The Cross-Modal Transformer(Pradeepkumar et al., 2024) is a Transformer-based method for classifying sleep stages. The proposed cross-modal Transformer consists of a cross-modal Transformer encoder architecture and a multi-scale one-dimensional convolutional neural network for automatic representation learning.

AttnSleep(Eldele et al., 2021) is composed of a multi-resolution convolutional neural network (MRCNN), an adaptive feature recalibration (AFR) module, and a temporal context encoder (TCE).

MMASleepNet(Zheng et al., 2022) is a model focused on extracting features from multimodal data, consisting of a multi-branch feature extraction (MBFE) module, followed by an attention-based feature fusion (AFF) module.

SleepPrintNet(Jia et al., 2020) consists of an EEG temporal feature extraction module, an EEG spectral-spatial feature extraction module for the time-frequency-space representation of EEG signals, and two multimodal feature extraction modules.

Comparison with the State-of-the-art Baselines

To evaluate the performance of the MGSleepNet model, we used a variety of state-of-the-art models to compare their performance against the overall results of the Sleep-edf dataset. The experimental results are shown in Table 1, 2, 3 and 4, where our model was compared with six baseline methods on the public dataset. The results indicate that, compared to the baseline models, our model achieved the best accuracy in overall performance while balancing effectiveness.

Table 1: Comparison of MGSleepNet's overall results with other models on the sleep-edf-20 dataset.

model	Overall Results(%)				
	ACC	MF1	Kappa	Sen	Pre
U-time	80.52	72.07	71.87	**77.86**	70.56
ResnetLSTM	81.96	**73.60**	**73.50**	75.51	**72.73**
Cross-Modal Transformer	81.29	72.49	72.73	76.12	70.60
AttnSleep	81.33	72.44	72.66	75.59	70.51
MMASleepNet	79.04	71.87	70.02	78.33	70.55
SleepPrintNet	81.57	73.27	72.97	76.51	71.71
MGSleepNet	**83.16**	73.04	75.33	74.01	72.54

Table 2: Comparison of MGSleepNet's overall results with other models on the sleep-edf-78 dataset.

model	Overall Results(%)				
	ACC	MF1	Kappa	Sen	Pre
U-time	79.74	71.84	70.97	73.50	69.78

ResnetLSTM	79.18	71.50	70.87	76.57	69.12
Cross-Modal Transformer	79.16	71.84	70.97	77.82	69.50
AttnSleep	80.06	73.62	72.27	**79.49**	70.72
MMASleepNet	76.62	69.53	67.92	78.05	67.39
SleepPrintNet	79.37	71.35	71.13	78.29	68.81
MGSleepNet	**82.46**	**72.72**	**73.27**	74.06	**72.03**

Ablation Study

To analyze the effectiveness of each module in MGSleepNet, we conducted an ablation study on the Sleep-edf dataset, with experimental results shown in Tables 3 and 4. In the ablation experiments, the GFI module was tested without the FIC module, while all other factors remained the same; the same applies to the FIC module. From the results, we conclude that using only the GFI module can achieve relatively good results, as it primarily extracts frequency, channel, and spatial information. However, adding the FIC module with fine-grained features significantly enhances the feature representation capability, resulting in better performance.

Table 3: Comparison of ablation results of MGSleepNet on the sleep-edf-20 dataset.

model	Overall Results(%)				
	ACC	MF1	Kappa	Sen	Pre
GFI	80.67	72.33	71.86	75.25	70.75
FIC	61.98	52.77	45.17	56.97	52.10
MGSleepNet	**83.16**	**73.04**	**75.33**	**74.01**	**72.54**

Table 4: Comparison of ablation results of MGSleepNet on the sleep-edf-78 dataset.

model	Overall Results(%)				
	ACC	MF1	Kappa	Sen	Pre
GFI	77.80	70.24	68.75	**74.59**	68.31
FIC	60.04	50.68	44.81	59.28	49.52
MGSleepNet	**82.46**	**72.72**	**73.27**	74.06	**72.03**

Test results of lab-built datasets

In order to further verify the practical application effects and generalization ability of the MGSleepNet model, we conducted additional experiments on a privately constructed sleep dataset in our laboratory. This dataset contains nighttime sleep records from multiple subjects. We employed the same training and evaluation strategies; however, due to the limited number of subjects being six, we used cross-subject six-fold cross-validation. The results obtained were Accuracy at 59.83%, Precision at 41.42%, Recall at 36.77%, MF1 Score at 34.90%, and Cohen Kappa at 27.06%. These results differ significantly from those of publicly available datasets, which may be attributed to the small number of subjects in our dataset. Through in-depth analysis of the private dataset, we found that MGSleepNet is capable of more effectively extracting and utilizing cross-modal information, which is crucial for improving the accuracy of sleep staging. Furthermore, MGSleepNet has shown excellent performance when faced with highly heterogeneous data, indicating its good adaptability and scalability, making it suitable for more diverse application scenarios.

Conclusion

This paper designs a multi-granularity sleep staging network capable of effectively extracting multimodal information from electroencephalograms (EEG) and electrooculograms (EOG). Its main components include the Global Feature Integration (GFI) module, which captures global features, and the Feature Integration and Classification (FIC) module, which extracts local features. Experimental results indicate that our model achieves state-of-the-art performance, thereby validating the complementarity of EEG and EOG information. Additionally, we conducted ablation experiments to further analyze the role of each module in our model, and we also tested on a real private dataset. In summary, MGSleepNet offers improved sleep staging performance using raw EEG and EOG signals without the need for additional data preprocessing. In future work, we will further attempt to optimize the model to enhance its generalization performance, making it more valuable for practical applications.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62171074, W2411084; the Chongqing Natural Science Fund Innovation and Development Joint Fund (Municipal Education Commission) project under Grant 2024NSCQ-LZX0058 and the funding of Institute for Advanced Sciences of Chongqing University of Posts and Telecommunications under Grant E011A2022327.

References

Bao, J., Wang, G., Wang, T., Wu, N., Hu, S.,& Lee, W.(2024). A Feature Fusion Model Based on Temporal Convolutional Network for Automatic Sleep Staging Using Single-Channel EEG. *IEEE Journal of Biomedical and Health Informatics*,28,(pp.6641-6652).

Lee, M., Kang, H., Yu, S., Cho, J., Oh, J.,& Lan, G.(2024). Automatic Sleep Stage Classification Using Nasal Pressure Decoding Based on a Multi-Kernel Convolutional BiLSTM Network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*,32,(pp.2533-2544).

Li, W., Liu, T., Xu, B., & Song A.(2024). ASleepFC: Feature Pyramid and Cross-Scale Context Learning for Sleep Staging.*IEEE Transactions on Neural Systems and Rehabilitation Engineering*,32,(pp.2198-2208)

Pradeepkumar, J., Anandakumar, M., Kugathasan, V., Suntharalingham, D., Kappel, S.,& Silva, A.(2024). Towards interpretable sleep stage classification using cross-modal transformers. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*,32,(pp.2893-2904).

An, P., Yuan, Z., Zhao, J., Jiang, X., & Du, B.(2021). An effective multi-model fusion method for EEG-based sleep stage classification. *Knowledge-Based Systems*,219,106890.

McCloskey, S., Jeffries, B., Koprinska, I., Miller, C., & Grunstein, R.(2019). An effective multi-model fusion method for EEG-based sleep stage classification. *Knowledge-Based Systems*,183, 104863.

Lin, Y., Wang, M., Hu, F., Cheng, X., & Xu, J.(2023). Multimodal Polysomnography-Based Automatic Sleep Stage Classification via Multiview Fusion Network. *IEEE Transactions on Instrumentation and Measurement*,73,(pp.1-12).

Jia, Z., Cai, X., &Jiao, Z.(2022). Multi-modal physiological signals based squeeze-and-excitation network with domain adversarial learning for sleep staging. *IEEE Sensors Journal*,22,(pp.3464-3471).

Yücelbaş, S., Yücelbaş, C., Tezel, G., Özsen, S., & Yosunkaya, S.(2024). Identification of full-night sleep parameters using morphological features of ECG signals: A practical alternative to EEG and EOG signals. *Biomedical Signal Processing and Control*,88,105633.

Kemp, B., Zwinderman, A-H., Tuk, B., Kamphuisen, H., & Oberye, J.(2000). Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Transactions on Biomedical Engineering* ,47,(pp.1185-1194).

Ary, L., Luis, A., Leon, G., Jeffrey, M.,JPlamen, C.,Ivanov, R.,Mark, J.,George, B.,Peng, C.,&Stanley, H.,(2000).PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 101,(pp.215-220).

Perslev, M., Jensen, M., Darkner, S., Jennum, P., & Igel, C.(2019). U-Time: A Fully Convolutional Network for Time Series Segmentation Applied to Sleep Staging. *Conference and Workshop on Neural Information Processing Systems* ,32.

Sun, Y., Wang, B., Jin, J., &Wang, X.,(2018).Deep Convolutional Network Method for Automatic Sleep Stage Classification Based on Neurophysiological Signals. *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, 101,(pp.1-5).

Pradeepkumar, J., Anandakumar M., Kugathasan, V., Suntharalingham, D., Kappel, S., Silva, A., & Edussooriya, C.(2024). Towards interpretable sleep stage classification using cross-modal transformers. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* ,32,(pp.2893-2904).

Eldele E., Chen Z., Liu , C., Wu, M., Kwoh, C., & Li, X.(2021). An attention-based deep learning approach for sleep stage classification with single-channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* ,29,(pp.809-818).

Zheng Y., Luo Y., Zou , B., Zhang, L., & Li, L.(2022). MMASleepNet: A multimodal attention network based on electrophysiological signals for automatic sleep staging. *Frontiers in Neuroscience* ,16,973761.

Jia Z., Cai X., Zheng , G., Wang, J., & Lin, Y.(2020). SleepPrintNet: A Multivariate Multimodal Neural Network Based on Physiological Time-Series for Automatic Sleep Staging. *IEEE Transactions on Artificial Intelligence* ,1,(pp.248-257).