2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

MHASSNet: A Deep Neural Network-based Automatic Sleep Staging Model Using Hybrid Attention Mechanism and State Space Model

Zhentao Huang[1], Shanwen Zhang[3] and 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
[3] Xi'an Key Laboratory of High Precision Industrial Intelligent Vision Measurement Technology, School of Electronic Information, Xijing University, Xi'an 710123, China
✉Corresponding author:tianyin@cqupt.edu.cn

Abstract. Automatic classification of sleep stages is an important method for assessing sleep quality. This paper introduces a sleep staging network (MHASSNet) that combines multi-scale convolution, hybrid attention mechanism, and state-space model, aiming to improve the accuracy of sleep stage classification for more effective evaluation of sleep quality. The model is characterized by its hybrid attention mechanism and multi-modal signal processing capabilities. First, MHASSNet uses a multi-scale convolutional neural network (MSCNN) to extract low-frequency and high-frequency features. Next, it employs a hybrid attention module (MAM) that integrates spatial and channel attention mechanisms to capture the significant spatiotemporal dependencies between these features. Additionally, the state-space model (SSM) is used to enhance the understanding of temporal contextual information. Experimental results show that when tested on two public datasets, MHASSNet achieves significant results across various evaluation metrics, demonstrating its superior performance and application potential in automatic sleep stage classification.

Keywords: Sleep Stage, Multi Scale Convolution, Attention Mechanism

1 Introduction

Sleep accounts for about one-third of a person's life and plays a crucial role in maintaining overall human health. With the acceleration of modern life rhythms and changes in lifestyle, various sleep disorders have become significant medical and public health issues, receiving increasing attention [1]. Sleep disorders can severely affect an individual's cognitive abilities, attention span, and mental state. Chronic sleep disturbances can even lead to hypertension, coronary artery disease, and other cardiovascular and cerebrovascular diseases [2]. Therefore, identifying and addressing patients' sleep disorders to improve sleep quality and ensure overall physical and mental health is critical.

Currently, sleep monitoring, sleep structure analysis, and sleep quality assessment have become research hotspots, particularly sleep staging, which is essential for evaluating sleep structure and diagnosing sleep-related disorders.In the field of sleep medicine, polysomnography (PSG) monitors are the 'gold standard' for diagnosing sleep disorders [3]. Polysomnography (PSG) is an important tool used to record various physiological signals, including electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG), playing a key role in sleep research [4-5]. In traditional methods, sleep specialists or doctors first segment the PSG data collected overnight into 30-second non-overlapping segments, and then manually analyze these segments according to authoritative standards (such as the Rechtschaffen and Kales standards (R&K) [6] or the American Academy of Sleep Medicine (AASM) standards [7]), categorizing them into different sleep stages. According to the AASM standards, sleep stages are typically divided into five categories: Wake (W), Non-Rapid Eye Movement 1 (N1), Non-Rapid Eye Movement 2 (N2), Non-Rapid Eye Movement 3 (N3), and Rapid Eye Movement Sleep (REM). Sleep specialists typically use these standards to manually classify the sleep stages, which is not only a labor-intensive process but also prone to subjective bias [8]. Therefore, automatic sleep staging is a more effective alternative to manual methods, with greater clinical value [9].

Simple deep learning networks cannot capture the time-varying features and time series information in sleep signals. To overcome these limitations, Supratak et al. [10] proposed a deep learning model, DeepSleepNet, which uses convolutional neural networks to extract time-invariant features and utilizes bidirectional long short-term memory to automatically learn the transition rules between sleep stages from EEG epochs. However, it primarily employs a one-to-one input-output model, aiming to label a single target sleep epoch at a time, but neglects the transition rules between different sleep stages. To address this issue, Phan et al. [11] proposed the SeqSleepNet hierarchical recurrent neural network, which treats the task as a sequence-to-sequence classification problem, receiving sequences of multiple epochs as input and classifying all labels at once. Due to the limited performance of single-channel EEG signal methods, in contrast, multi-channel sleep graphs significantly improve accuracy with their multi-channel structure and enhance the interpretability of the sleep stage classification model results. For example, Cheng et al. [12] proposed a new distributed multimodal and multi-label decision system (MML-DMS). It includes several interconnected classifier modules, including deep convolutional neural networks (CNN) and shallow perceptron neural networks (NN). Chambon et al. [13] proposed a neural network model that uses multi-channel, multi-modal signals as input data. These methods mainly leverage multi-channel features in PSG signals and combine the features of different channels in a cascading manner. Lin et al. [14] proposed a multi-scale local feature extractor (MSLFE), which has a multi-branch convolutional neural network (CNN) with different convolution kernel sizes and a global relationship modeling (GRM) module, to effectively extract features in both the time and frequency domains. Duan et al. [15] proposed MMS-SleepNet, which uses a deep learning multi-modal feature extraction module (MMS-FE) to embed expert knowledge, effectively capturing the multi-modal features of each stage and the fine-grained EEG features of different frequencies. However, these approaches fail to capture the cross-modal contextual relationships between different modalities, overlooking the differences in how each modality contributes to sleep stages, which affects the overall performance of the sleep model.In summary, the existing research has the following shortcomings: (1) it is difficult to extract effective features from the raw EEG signals; (2) some studies focus on extracting channel features from the signals while neglecting local features, which impacts the overall performance and generalization ability of the model; (3) most models do not delve deeply enough into the complex relationships in cross-modal contexts. Although some studies have fused multimodal data through concatenation, addition, or dot multiplication [14-16], these simple methods fail to fully integrate global contextual information and shared characteristics between different samples. To address these research deficiencies, this paper proposes an end-to-end multimodal automatic sleep staging network model based on EEG and EOG. First, a multi-scale 1D convolutional neural network (1D-CNN) is used to extract features corresponding to low and high frequencies from different signal bands. Furthermore, a hybrid attention mechanism (MAM) that combines local spatial information and global channel information is then incorporated. Finally, a state-space sequence coupling module is introduced to learn the cross-modal contextual relationships between the signals. This paper also conducts extensive experiments on the sleep-EDF dataset to comprehensively verify the effectiveness and feasibility of the proposed method. Through in-depth testing on this dataset, its reliability and applicability are further confirmed.

2 Methodology

2.1 The Structure of MHASSNet

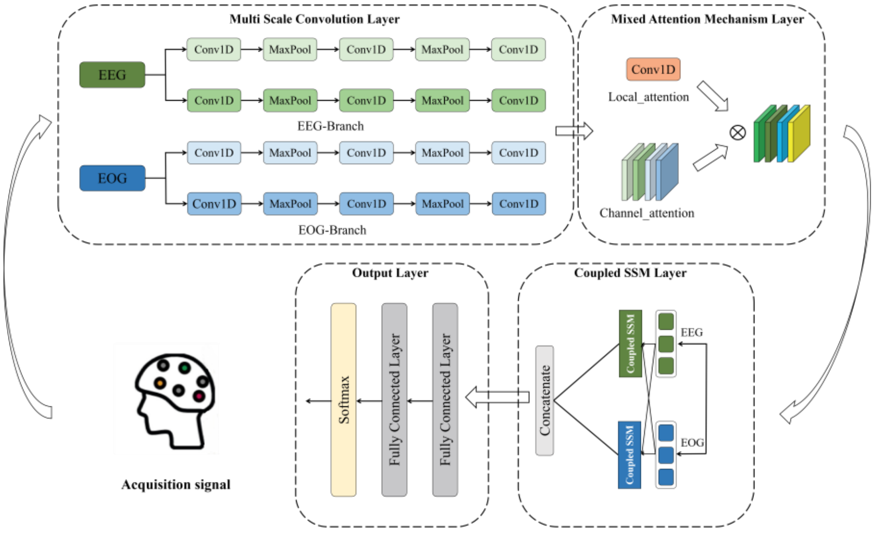

Fig. 1. MHASSNet framework

As shown in Figure 1, the MHASSNet architecture mainly includes a multi-branch feature extraction module, an attention-based feature fusion module, and a classification module. The multi-branch feature extraction module extracts different frequency signal features from EEG and EOG using convolutional kernels of two different sizes. The attention-based feature fusion module is primarily composed of a Mixed Attention mechanism layer and a coupled SSM layer. The Mixed Attention mechanism layer combines local spatial information with global channel information, evaluating and learning the importance of each channel in the feature map. These weights are then applied to the original channel data, enhancing the expressiveness of key features. The coupled state-space model layer focuses on extracting temporal features from multimodal sleep signals, improving the model's understanding by deeply exploring cross-modal temporal dynamic relationships. Finally, the classification module's fully connected layer integrates these features and outputs the classification results. This model performs an end-to-end analysis of sleep signal features from both temporal and spatial perspectives, improving the accuracy of the prediction results.

2.2 Multiscale Convolution

For the refined extraction of the spatiotemporal features of EEG signals, numerous studies have confirmed [16-17] that the Multi-Scale Convolutional Neural Network (MSCNN) architecture demonstrates unique advantages in heterogeneous data processing. MSCNN optimizes feature extraction by constructing a dual-branch pathway, which separately targets the low-frequency and high-frequency features in the signal for efficient capture. In terms of specific implementation, MSCNN adopts a parallel dual-scale convolution structure. For example, in the first layer, the large-scale convolution kernel is (1, 50), with a stride of 20 and 64 kernels, achieving robust extraction of low-frequency features through a wide receptive field; while the small-scale CNN convolution kernel is (1, 20), with a stride of 5 and 64 kernels, focusing on the high-frequency detailed features in the signal. Additionally, Maxpooling(4, 4) refers to a maximum pooling layer with a kernel size of (4, 4). After multiple layers of convolution, EEG and EOG are connected together via the Concatenation operation and then output to the next layer of the network.

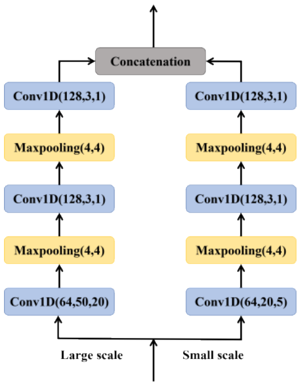

Fig. 2. MHASSNet framework

2.3 Mixed Attention Mechanism

The attention mechanism is a data processing technique and one of the core capabilities of human perception and decision-making. In brainwave signal processing applications, commonly used attention mechanisms include Squeeze-and-Excitation Attention and Efficient Channel Attention, among other channel attention mechanisms. However, these only consider the overall relationship between channels, neglecting the spatial information of each channel. Therefore, for the multimodal signals discussed in this paper, as shown in Figure 3, this paper introduces a hybrid attention mechanism that combines local spatial information with global channel information, significantly enhancing the model's overall performance in feature extraction, expression capability, and computational efficiency.

First, the local attention mechanism uses a one-dimensional convolution operation to effectively capture short-range dependencies on the time series, enhancing the model's sensitivity to local patterns. The one-dimensional convolution operation computes the local attention mechanism as shown in formula (1). Let the input tensor be x∈$R^{B\times C\times T}$, where B is the batch size, C is the number of channels, and T is the time step. The local attention mechanism is then applied, and the output has C/r channels, a convolution kernel size of k, with padding set to k/2, where r represents the reduction ratio and k represents the convolution kernel size. The output is local_attention.

$$local\_attention = \mathrm{Re}\,LU(Conv1D(X)) \tag{1}$$

Meanwhile, as shown in equation (2-3), the channel attention mechanism is used. The channel attention first computes the time step average for each channel and generates the channel attention weights through a fully connected layer:FC is a linear transformation, with an input dimension of C and an output dimension of C/r.

$$z_c = \frac{1}{T}\sum_{t=1}^{T} x[:,c,t], c = 1,2,...,C \tag{2}$$

$$channel\_attention = \mathrm{Re}\,LU(FC(z)) \tag{3}$$

Finally, as shown in formula (4), the local attention and channel attention are multiplied, where ⊙ denotes element-wise multiplication.

$$\text{combined}\_attention = \text{local}\_attention \otimes \text{channel\_}attention \tag{4}$$

The combination of both not only captures features across multiple scales but also significantly reduces computational complexity and parameter count through dimensionality reduction (reduction ratio), improving computational efficiency while maintaining performance. Additionally, this mechanism enhances the model's ability to adapt to complex input distributions through non-linear activation functions (such as ReLU) and element-wise multiplication, demonstrating stronger robustness.

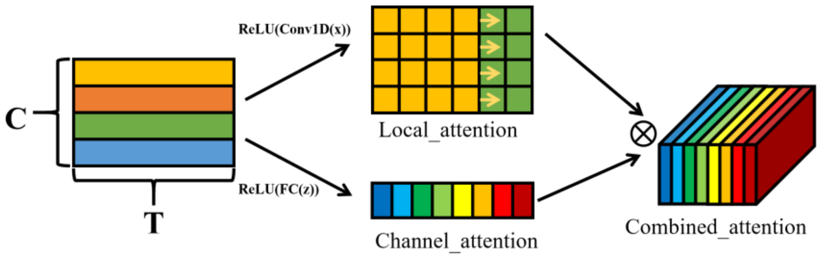

Fig. 3. Mixed attention mechanism

2.4 State Space Model Coupling Module

Recently, the State Space Model (SSM) has gradually emerged as an efficient building unit or hierarchical structure in the process of constructing deep networks. However, SSM is unable to capture the interaction information between modalities. Therefore, this paper proposes a multi-modal SSM coupling module that combines SSM with the attention mechanism, aiming to enhance the interaction ability between modalities.The SSM coupling module is shown in Figure 4. The data processing flow of the SSM model can be described as follows: First, the modality input is defined as XnϵR$^{B\times T\times D}$, where B represents the batch size, T is the time step, and D is the dimension of input features. Additionally, the model includes the following key matrices: state update matrix AϵR$^{H\times H}$, input projection matrix BϵR$^{H\times D}$, modality correlation matrix SnϵR$^{H\times H}$, and output transformation matrix EϵR$^{D\times H}$. Here, H is the hidden state dimension, and we also introduce a weighted fusion parameter Wn to adaptively assess the importance of each modality.

State update mechanism: First, according to equation (5), layer normalization is applied to the input of the nth modality at time step t to standardize the input features. Then, based on equation (6), a cross-modal interaction representation is constructed by aggregating the hidden states of all modalities.

$$\tilde{x}_n(t) = LayerNorm(x_n(t)) \tag{5}$$

$$H_{sum}(t) = \sum_{n=1}^{N} h_n(t-1) \tag{6}$$

Next, the hidden state for the next time step is generated by combining the input features at the current time step with the historical hidden states of each modality. Specifically, the hidden state update rule for the n-th modality is given by formula (7):

$$h_n(t) = \tanh\left(\sum_{m=1}^{M} h_m(t-1)\cdot S_n + X\tilde{(t)}\cdot B^T\right) \in R^{B\times H} \tag{7}$$

Among them, M represents the number of modes, h$_m$(t−1) is the hidden state of the m-th mode at time step t−1, Sn is the correlation matrix of mode n, and BT is the transpose of the input projection matrix.

Cross-modal interaction and attention mechanism: In order to further enhance the interaction effects between modalities, by combining the historical states of all modalities with the current input, according to formula (8), the similarity score score(h_n(t), h_m(t)) between modality n and modality m at time step t is calculated to achieve adaptive weighting between modalities. This score reflects the degree of correlation between the two modalities at the current time step. Finally, as shown in formula (9), the hidden state of modality n at time step t will interact with the states of other modalities through attention weighting:

$$\alpha_{nm}(t) = \frac{\exp(score(h_n(t), h_m(t)))}{\sum_{m=1}^{M} \exp(score(h_n(t), h_m(t)))} \tag{8}$$

$$h_n(t) = \tanh(\sum_{m=1}^{M} \alpha_{nm}(t) \cdot h_n(t-1) \cdot S_n + \tilde{X}(t) \cdot B^T) \tag{9}$$

Output: According to formula (10), the output Yn of each modality is obtained by projecting the corresponding hidden state h_n(t) into the output space through a linear transformation matrix E^T. Then, according to formula (11), the model performs a weighted integration of the outputs from all modalities to generate the final multimodal output Y. Here, W_n is the weighted parameter for each modality.

$$Y_n = h_n \cdot E^T \tag{10}$$

$$Y = \frac{1}{M} \sum_{n=1}^{M} W_n \cdot Y_n \tag{11}$$

3 Materials and experimental setup

3.1 Dataset

In this paper, two public datasets, Sleep-EDF-20 and Sleep-EDF-78, were used from Physiobank [19]. Table 1 provides detailed information, including the number of samples and the sample proportions in the datasets. Sleep-EDF-20 contains sleep PSG data files from 20 subjects, while Sleep-EDF-78 contains data from 78 subjects. The data collected from the subjects includes EEG signals from two channels (from the Fpz-Cz and Pz-Oz electrode positions), an EOG signal (horizontal), an EMG signal (chin), and event markers. Additionally, both the EOG and EEG signals were sampled at 100 Hz. The entire night of data is recorded in two files: SC*PSG. The hypnogram was manually scored according to the Rechtschaffen and Kales protocol and then labeled as N1, N2, N3, N4, Wake, REM, and UNKNOWN classes.

Table 1. Sample Count and Sample Proportion in the Sleep-EDF Dataset.

Dataset	W	N1	N2	N3	REM	total
Sleep-EDF-20	8285	2804	17799	5703	7717	42308
	19.6%	6.6%	42.1%	13.5%	18.2%	
Sleep-EDF-78	65951	21522	69132	13039	25835	195479
	14.3%	3.2%	43.7%	18.5%	20.3%	

3.2 Contrast model

The proposed model is compared with the following baseline methods:U-time [20] is a time-based fully convolutional network based on the U-Net architecture. It maps input sequences of arbitrary length to class label sequences on a freely chosen time scale. ResnetLSTM [21] utilizes residual modules to increase the network depth for extracting multi-level features of sleep staging, while using Long Short-Term Memory (LSTM) networks to learn the sleep transition mechanisms during the sleep process. The Cross-Modal Transformer [22] proposes an architecture consisting of a cross-modal Transformer encoder and a multi-scale 1D convolutional neural network for automatic representation learning. AttnSleep [23] is based on a multi-resolution convolutional neural network (MRCNN), an adaptive feature recalibration (AFR) module, and a temporal context encoder (TCE). MMASleepNet[24] is a model that extracts multimodal features. It has a multi-branch feature extraction (MBFE) module, followed by an attention-based feature fusion (AFF) module.

3.3 Model parameter setting

In this experiment, to evaluate the model's performance, we applied a five-fold cross-subject method to two datasets. Specifically, participants in each dataset were randomly but evenly divided into five groups to ensure that each fold reflects the overall sample's characteristic distribution. During the model training phase, we used the Adam optimizer with a learning rate of 0.001, and to address the class imbalance issue in the dataset, we adopted a weighted cross-entropy loss function to adjust the importance weights of different class samples, ensuring the model could learn information from each class in a balanced manner. Furthermore, the batch size was set to 128 during training, and the entire training process was executed over 100 epochs, aiming to enhance the model's expressiveness and generalization ability through sufficient iteration and optimization.

3.4 Evaluation indexes

To evaluate the model's performance, we used a set of commonly used evaluation metrics. The formulas for the evaluation metrics such as accuracy, Macro-F1 score, Cohen's Kappa, sensitivity, specificity, and precision are shown in (12)-(17).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Macro\text{-}F1\ score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} \tag{13}$$

$$Cohen'Kappa = \frac{p_o - p_e}{1 - p_e} \tag{14}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{15}$$

$$Specificity = \frac{TN}{TN + FP} \tag{16}$$

$$precision = \frac{TP}{TP + FP} \tag{17}$$

4 Experimental Results and Analysis

4.1 Comparison with the State-of-the-art Baselines

Tables 2 and 3 show a comparison of six models: U-time, ResnetLSTM, Cross-Modal Transformer, AttnSleep, MMASleepNet, and MHASSNet. Among them, MMASleepNet has the worst performance.MMASleepNe on the Sleep-EDF-20 Dataset achieved an Accuracy of 79.04%, Macro-F1 score of 71.87%, Cohen's Kappa of 70.02%, sensitivity of 78.33%, specificity of 94.81%, and Precision of 70.55%. On the Sleep-EDF-78 Dataset, the Accuracy was 76.62%, Macro-F1 score 69.53%, Cohen's Kappa 67.92%, sensitivity 78.05%, specificity 94.26%, and Precision 67.39%. The MHASSNet proposed in this paper achieved the following results on the Sleep-EDF-20 Dataset: Accuracy 84.29%, Macro-F1 score 73.44%, Cohen's Kappa 75.96%, sensitivity 71.94%, specificity 95.31%, and Precision 76.29%. On the Sleep-EDF-78 Dataset, the results were: Accuracy 82.13%, Macro-F1 score 70.52%, Cohen's Kappa 73.78%, sensitivity 69.14%, specificity 94.90%, and Precision 74.37%. Furthermore, as can be seen from the table, the classification accuracy of N2, N3, and Wake is relatively high for both the Sleep-EDF-20 and Sleep-EDF-78 datasets. The MHASSNet proposed in this paper outperforms the baseline methods, indicating that MHASSNet has an advantage in automatic sleep staging during feature extraction and fusion operations of multimodal electrophysiological signals.

Table 2. Performance of six models on the Sleep-EDF-20 dataset.

model	Per-Class Precisions(%)						Per-Class Precisions(%)				
	ACC	MF1	Kappa	Sen	Spec	Pre	Pre(N1)	Pre(N2)	Pre(N3)	Pre(REW)	Pre(Wake)
U-time	80.52	72.07	71.87	77.86	95.03	70.56	37.66	92.99	57.24	75.89	89.02
ResnetLSTM	81.96	73.60	73.50	75.51	95.15	72.73	37.92	90.77	68.87	76.12	89.97
Cross-Modal Transformer	81.29	72.49	72.73	76.12	95.13	70.60	36.76	92.72	61.76	73.82	87.94
AttnSleep	81.33	72.44	72.66	75.59	95.09	70.51	36.02	92.81	64.42	73.53	85.70
MMASleepNet	79.04	71.87	70.02	78.33	94.81	70.55	33.36	94.24	57.16	77.13	90.83
MHASSNet	84.29	73.44	75.96	71.94	95.31	76.29	46.54	88.63	79.91	75.93	90.47

Table 3. Performance of six models on the Sleep-EDF-78 dataset.

model	Per-Class Precisions(%)						Per-Class Precisions(%)				
	ACC	MF1	Kappa	Sen	Spec	Pre	Pre(N1)	Pre(N2)	Pre(N3)	Pre(REW)	Pre(Wake)
U-time	79.74	71.84	70.97	73.50	94.79	69.78	38.97	84.42	60.04	72.94	92.53
ResnetLSTM	79.18	71.50	70.87	76.57	94.75	69.12	40.15	88.86	52.08	70.48	941.8
Cross-Modal Transformer	79.16	71.84	70.97	77.82	94.79	69.50	39.26	89.52	50.15	74.63	93.92
AttnSleep	80.06	73.62	72.27	79.49	95.01	70.72	42.00	89.16	53.91	73.25	95.29
MMASleepNet	76.62	69.53	67.92	78.05	94.26	67.39	38.57	88.37	41.09	73.90	95.04
MHASSNet	82.13	70.52	73.78	69.14	94.90	74.37	46.41	83.82	75.14	77.67	88.79

4.2 Ablation Study

MMASleepNet consists of MSCNN module, MAM module, and SSM module. To analyze the impact of each module and validate the effectiveness of the modes in MHASSNet, ablation experiments were designed on the Sleep-EDF-20 and Sleep-EDF-78 datasets as follows. Since the MSCNN and MAM modules are spatial feature extraction modules, while SSM is a temporal feature extraction module, this paper mainly compares the impact of these two types of modules on the overall performance of the model. From Table 4 and Table 5, it can be seen that the performance of the MSCNN-MAM combined model is slightly lower than the overall performance of the MHASSNet model. This indicates that the MSCNN-MAM module plays a major role in the overall model. Additionally, by adding the SSM module, which is capable of extracting temporal features, the classification performance can be further improved, proving the necessity of modeling the interdependencies between features.

Table 4. Results of MHASSNe's ablation experiment on the Sleep-EDF-20 dataset.

model	Per-Class Precisions(%)						Per-Class Precisions(%)				
	ACC	MF1	Kappa	Sen	Spec	Pre	Pre(N1)	Pre(N2)	Pre(N3)	Pre(REW)	Pre(Wake)
MSCNN-MAM	82.29	69.55	72.80	68.64	94.67	73.85	39.25	87.47	78.91	76.19	87.41
SSM	51.60	22.24	11.39	24.76	82.18	29.31	0	55.08	23.64	33.79	34.02
MHASSNet	84.29	73.44	75.96	71.94	95.31	76.29	46.54	88.63	79.91	75.93	90.47

Table 5. Results of MHASSNe's ablation experiment on the Sleep-EDF-78 dataset.

model	Per-Class Precisions(%)						Per-Class Precisions(%)				
	ACC	MF1	Kappa	Sen	Spec	Pre	Pre(N1)	Pre(N2)	Pre(N3)	Pre(REW)	Pre(Wake)
MSCNN-MAM	82.29	69.55	72.80	68.64	94.67	73.85	39.25	87.47	78.91	76.19	87.41
SSM	51.60	22.24	11.39	24.76	82.18	29.31	0	55.08	23.64	33.79	34.02
MHASSNet	84.29	73.44	75.96	71.94	95.31	76.29	46.54	88.63	79.91	75.93	90.47

5 Conclusion

This paper introduces a multi-channel automatic sleep signal classification method, namely the MHASSNet model. The model uses multi-scale CNNs to extract features from different frequency bands of EEG signals, and then employs a Mixed Attention Mechanism module to extract the feature weights of channels and spatial aspects. It also incorporates the SSM module to capture global context features for each sample, and finally outputs sleep stage results. The proposed MHASSNet model is an end-to-end spatiotemporal feature fusion network, achieving accuracy rates of 84.29% and 82.13% in five-fold cross-subject classification tasks on the Sleep-EDF-20 and Sleep-EDF-78 datasets, demonstrating the outstanding capability of the proposed deep learning model. Furthermore, the paper also analyzes the effectiveness of each module's fusion in the MHASSNet model through ablation experiments. This model holds promise for designing efficient and accurate real-time brain-machine interface frameworks.

Acknowledgments. This work was supported by the Chongqing Natural Science Fund Innovation and Development Joint Fund (Municipal Education Commission) project under Grant CSTB2024NSCQ-LZX0121; and the funding of Institute for Advanced Sciences of Chongqing University of Posts and Telecommunications under Grant E011A2022327.

Disclosure of Interests. authors have no competing interests.

References

1. Ramar, K., Malhotra, R. K., Carden, K. A., Martin, J. L., Abbasi-Feinberg, F., Aurora, R. N., ... & Trotti, L. M. Sleep is essential to health: an American Academy of Sleep Medicine position statement. Journal of Clinical Sleep Medicine, 17(10), 2115-2119(2021)

2. Moon, C., Phelan, C. H., Lauver, D. R., & Bratzke, L. C. A narrative review of how sleep-related breathing disorders and cardiovascular diseases are linked: An update for advanced practice registered nurses. Clinical Nurse Specialist, 30(6), 347-362(2016)
3. Shambroom, J. R., Fábregas, S. E., & Johnstone, J. Validation of an automated wireless system to monitor sleep in healthy adults. Journal of sleep research, 21(2), 221-230(2012)
4. Zhao, R., Xia, Y., & Wang, Q. Dual-modal and multi-scale deep neural networks for sleep staging using EEG and ECG signals. Biomedical Signal Processing and Control, 66, 102455(2021)
5. Chen, J., Han, Z., Qiao, H., Li, C., & Peng, H. EEG-based sleep staging via self-attention based capsule network with Bi-LSTM model. Biomedical Signal Processing and Control, 86, 105351(2023)
6. Wolpert, E. A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Archives of General Psychiatry, 20(2), 246-247(1969)
7. Iber C. The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specification[J], 2007.
8. Chen, X., Chen, Y., Ma, W., Fan, X., & Li, Y. Toward sleep apnea detection with light-weight multi-scaled fusion network. Knowledge-Based Systems, 247, 108783 (2022).
9. Acosta, J. N., Falcone, G. J., Rajpurkar, P., & Topol, E. J. Multimodal biomedical AI. Nature medicine, 28(9), 1773-1784(2022).
10. Supratak, A., Dong, H., Wu, C., & Guo, Y. DeepSleepNet: A model for automatic sleep stage s`ing based on raw single-channel EEG. IEEE transactions on neural systems and rehabilitation engineering, 25(11), 1998-2008(2017).
11. Phan, H., Andreotti, F., Cooray, N., Chén, O. Y., & De Vos, M. SeqSleepNet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(3), 400-410(2019).
12. Cheng, Y. H., Lech, M., & Wilkinson, R. H. Simultaneous sleep stage and sleep disorder detection from multimodal sensors using deep learning. Sensors, 23(7), 3468(2023).
13. 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 Transactions on Neural Systems and Rehabilitation Engineering, 26(4), 758-769(2018).
14. Lin, Y., Wang, M., Hu, F., Cheng, X., & Xu, J. Multimodal polysomnography-based automatic sleep stage classification via multiview fusion network. IEEE Transactions on Instrumentation and Measurement, 73, 1-12(2023).
15. Duan, L., Ma, B., Yin, Y., Huang, Z., & Qiao, Y. MMS-SleepNet: A knowledge-based multimodal and multiscale network for sleep staging. Biomedical Signal Processing and Control, 103, 107370(2025).
16. Liu, K., Xing, X., Yang, T., Yu, Z., Xiao, B., Wang, G., & Wu, W. DMSACNN: Deep Multiscale Attentional Convolutional Neural Network for EEG-Based Motor Decoding. IEEE Journal of Biomedical and Health Informatics.(2025).
17. Wu, X., Chu, Y., Li, Q., Luo, Y., Zhao, Y., & Zhao, X AMEEGNet: attention-based multiscale EEGNet for effective motor imagery EEG decoding. Frontiers in Neurorobotics, 19, 1540033(2025)..
18. Wan, D., Lu, R., Shen, S., Xu, T., Lang, X., & Ren, Z Mixed local channel attention for object detection. Engineering Applications of Artificial Intelligence, 123, 106442(2023).
19. Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A., & Oberye, J. J Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. IEEE Transactions on Biomedical Engineering, 47(9), 1185-1194(2000)..

20. Perslev, M., Jensen, M., Darkner, S., Jennum, P. J., & Igel, C. U-time: A fully convolutional network for time series segmentation applied to sleep staging. Advances in neural information processing systems, 32(2019).
21. Sun, Y., Wang, B., Jin, J., & Wang, X. Deep convolutional network method for automatic sleep stage classification based on neurophysiological signals. In 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (pp. 1-5). IEEE (2018).
22. Pradeepkumar, J., Anandakumar, M., Kugathasan, V., Suntharalingham, D., Kappel, S. L., De Silva, A. C., & Edussooriya, C. U. Towards interpretable sleep stage classification using cross-modal transformers. IEEE Transactions on Neural Systems and Rehabilitation Engineering(2024).
23. Eldele, E., Chen, Z., Liu, C., Wu, M., Kwoh, C. K., Li, X., & Guan, C. An attention-based deep learning approach for sleep stage classification with single-channel EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29, 809-818(2021).
24. Yubo, Z., Yingying, L., Bing, Z., Lin, Z., & Lei, L. MMASleepNet: A multimodal attention network based on electrophysiological signals for automatic sleep staging. Frontiers in Neuroscience, 16, 973761(2022).