BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Towards interpretable sleep stage classification with a multi-stream fusion network

Jingrui Chen[1], Xiaomao Fan[2*], Ruiquan Ge[4], Jing Xiao[3], Ruxin Wang[5], Wenjun Ma[3*] and Ye Li[5]

Abstract

Sleep stage classification is a significant measure in assessing sleep quality and diagnosing sleep disorders. Many researchers have investigated automatic sleep stage classification methods and achieved promising results. However, these methods ignored the heterogeneous information fusion of the spatial–temporal and spectral–temporal features among multiple-channel sleep monitoring signals. In this study, we propose an interpretable multi-stream fusion network, named MSF-SleepNet, for sleep stage classification. Specifically, we employ Chebyshev graph convolution and temporal convolution to obtain the spatial–temporal features from body-topological information of sleep monitoring signals. Meanwhile, we utilize a short time Fourier transform and gated recurrent unit to learn the spectral–temporal features from sleep monitoring signals. After fusing the spatial–temporal and spectral–temporal features, we use a contrastive learning scheme to enhance the differences in feature patterns of sleep monitoring signals across various sleep stages. Finally, LIME is employed to improve the interpretability of MSF-SleepNet. Experimental results on ISRUC-S1 and ISRUC-S3 datasets show that MSF-SleepNet achieves competitive results and is superior to its state-of-the-art counterparts on most of performance metrics.

Keywords Sleep stage classification, Fusion network, Contrastive learning, Chebyshev graph convolution, Model interpretability

Introduction

Sleep is a critical physiological activity for human beings, occupying a third of a person's life span. Sleep quality is directly connected to physical and mental health; *e.g.*, low-quality sleep can lead to a variety of health problems, such as stroke, and brain damage [1, 2]. According to the American Academy of Sleep Medicine (AASM), an overnight sleep can be divided into three main stages: wake (W), rapid eye movement (REM), and non rapid eye movement (NREM) [3]. NREM can be further divided into N1, N2, and N3 stages. The determination of sleep stages is a widely used measure for physicians to evaluate sleep quality, as it aids them in accurately diagnosing the disease and formulating a reasonable treatment.

*Correspondence:
Xiaomao Fan
astrofan2008@gmail.com
Wenjun Ma
mawenjun@scnu.edu.cn
[1]Department of Information Management, Guangdong Justice Police Vocational College, Guangzhou, Guangdong 510520, China
[2]College of Big Data and Internet, Shenzhen Technology University, Shenzhen, Guangdong 518055, China
[3]School of Computer Science, South China Normal University, Guangzhou, Guangdong 510631, China
[4]School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China
[5]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong 518055, China

In clinical practice, polysomnography (PSG) is the gold standard for sleep stage classification, and mainly includes electroencephalogram (EEG) signals, electrooculogram (EOG) signals, electromyography (EMG) signals, and electrocardiogram (ECG) signals [4, 5]. Many researchers have attempted to investigate automatic sleep stage classification methods and achieved promising results. Generally, automatic sleep classification methods can be divided into conventional feature-based methods employing machine learning [6] and end-to-end-based methods using deep neural networks [7]. For feature-based methods, specified features are generated from sleep monitoring signals, and then are fed into machine learning models like those employing a k-nearest neighbors (k–NN) algorithm [8], support vector machine (SVM) [9], random forest (RF) [10], decision tree (DT) [11], naive Bayes classifier [12], or linear and quadratic discriminant analysis [13, 14]. However, feature-based methods highly rely on the quality of sleep monitoring signals and domain knowledge utilization, and this greatly limits their classification performance in real scenarios. Note that deep neural networks have achieved great success in computer vision, natural language processing, etc.; therefore, researchers have shifted their focus from feature-based methods to end-to-end-based methods, which have the advantage of effectively self-learning features from original sleep monitoring signals. Recent end-to-end-based methods are MNN [15], DeepSleepNet [16], SeqSleepNet [17], GraphSleepNet [18], MSTGCN [19], MVF-SleepNet [20], TinySleepNet [21], U-Sleep [22], and so on. These methods utilize the temporal, spectral, spatial–temporal, or spectral–temporal feature information extracted from sleep monitoring signals to build sleep classification models. However, they still ignore the heterogeneous information fusion of the spatial–temporal and spectral–temporal features information, which would enhance sleep classification performance.

In this study, we propose an interpretable multi-stream fusion network, named MSF-SleepNet, for sleep stage classification. Concretely, graph structure learning, Chebyshev graph convolution, a spatial and temporal attention mechanism, and temporal convolution are used to acquire the spatial–temporal features from body-topological information of sleep monitoring signals. Simultaneously, a short time Fourier transform (STFT), VGG-16 network, gated recurrent unit (GRU), and GRU attention are employed to capture the spectral–temporal features from sleep monitoring signals. Finally, we fuse the spatial–temporal and spectral–temporal features to classify sleep stages into five categories, and a contrastive learning scheme is utilized to improve the feature pattern difference of sleep monitoring signals between different sleep stages. LIME is applied to enhance the model's interpretability. Experiments are performed on two public datasets: ISRUC-S1 and ISRUC-S3 [23]. The results demonstrate that MSF-SleepNet achieves competitive results and outperforms its state-of-the-art counterparts on most of performance metrics.

To sum up, our contributions are as follows.

- We use the combination of Chebyshev graph convolution and temporal convolution to learn the spatial–temporal features from body-topological information of sleep monitoring signals.
- We combine STFT and GRU to obtain the spectral–temporal features from sleep monitoring signals.
- We fuse spatial–temporal and spectral–temporal features, and utilize a contrastive learning scheme to enhance the differences in feature patterns of sleep monitoring signals across various sleep stages.
- LIME is applied to improve the interpretability of MSF-SleepNet.
- Experiment results on ISRUC-S1 and ISRUC-S3 datasets reveal that MSF-SleepNet achieves state-of-the-art performance.

The rest of this paper is organized as follows. Section II provides a concrete introduction of existing sleep stage classification methods, their deficiencies, and some new models for dealing with these defects. Section III introduces all components of MSF-SleepNet in details. Two datasets of ISRUC-S1 and ISRUC-S3 are introduced to compare MSF-SleepNet with cutting-edge baselines, and the results of an interpretability analysis are presented in Section IV. The final Section presents our conclusions.

Related work

Feature-based methods employing machine learning

Various feature-based methods employing machine learning have been applied to automatic sleep stage classification. Machine learning classifiers can be grouped [24] as instance-based algorithms, decision tree algorithms, and Bayes rule-based classifiers. Memar et al. [25] proposed a system to classify the wake and sleep stages with high sensitivity and specificity. Then, the minimal-redundancy-maximal-relevance feature selection algorithm is used to eliminate redundant and irrelevant features. Finally, selected features are classified by a random forest classifier. Dhok et al. [26] proposed an automatic classification method for CAP phases (A and B) based on the Wigner–Ville Distribution (WVD) and Rényi entropy (RE) features. A support vector machine based on a medium Gaussian kernel is used for classification, and 10-fold cross validation is conducted. Gunnarsdottir et al. [27] developed an algorithm to extract features from EEG, EMG, and EOG signals using a likelihood ratio decision tree classifier based on rules

predefined in the AASM manual. Features are calculated on 30-second epochs in the time and frequency domain of the signal; these features are then input to the classifier. However, feature-based methods highly rely on the quality of sleep monitoring signals and feature extraction/selection with domain knowledge [24, 28].

End-to-end-based methods using deep neural networks

With the big success in computer vision, natural language process, and other domains, more and more end-to-end-based methods using deep neural networks are being utilized to classify sleep stages. These methods employed convolutional neural networks (CNNs) [29], recurrent neural networks (RNNs) [17], graph convolution networks (GCNs) [30], long short-term memory (LSTM) [31], and GRUs. Sokolovsky et al. [32] designed a deep CNN architecture for automated sleep stage classification of EEG and EOG signals. Dong et al. [15] used a mixed neural networks concatenating a multi-layer perceptron and an RNN to address sleep stage classification with a single electrode recording. Furthermore, Phan et al. [33] proposed a joint classification prediction framework based on a CNN for automatic sleep stage, and introduced a simple and efficient CNN architecture to support the framework. This framework provided a way to further study different automatic sleep stage neural networks. Jia et al. [19] proposed a multi-view spatial–temporal graph convolutional network (MSTGCN) with domain generalization for sleep stage classification. Although these end-to-end-based methods can effectively utilize time-varying spatial and temporal features from multi-channel brain signals, they still ignored the heterogeneous information fusion of the spatial–temporal and spectral–temporal features information, which would greatly improve the sleep staging performance.

Methodology

Figure 1 shows the overall structure of MSF-SleepNet. It contains three modules: a spatial–temporal feature extractor, spectral–temporal feature extractor, and sleep stage classifier. Details are presented below.

Spatial–temporal feature extractor

This module consists of sleep contrastive learning network (SCL Net) and the process of learning the spatial–temporal features.

Sleep contrastive learning network (SCL Net)

The SCL Net consists of a framework for contrastive learning of sleep stage classification (SleepCL) and a 1-D convolution network named SFNet.

Contrastive learning [34] is a kind of self-supervised learning that does not rely on annotated data and instead learns knowledge by itself from unannotated data. The idea of contrastive learning can be expressed as acquiring a representation learning model by automatically constructing similar and dissimilar instances. Through this model, similar instances are relatively close together in projection space, while dissimilar instances are relatively far apart [35]. Hinton et al. proposed the SimCLR [36] framework in 2020; it consists of data augmentation, a base encoder, a projection head, a function for maximizing similarity, and a contrastive loss function.

In our network, we applied the idea of contrastive learning with the SimCLR framework to enhance feature representation extraction. Specifically, we combine SleepCL with SFNet to acquire general feature representations from unlabeled data. The resulting contrastive learning network is displayed in Fig. 2. We use minimum–maximum normalization [37] and Z-score normalization [38] to transform original signal data into augmented data. For the base encoder, we design a 1-D convolution neural network named SFNet to extract

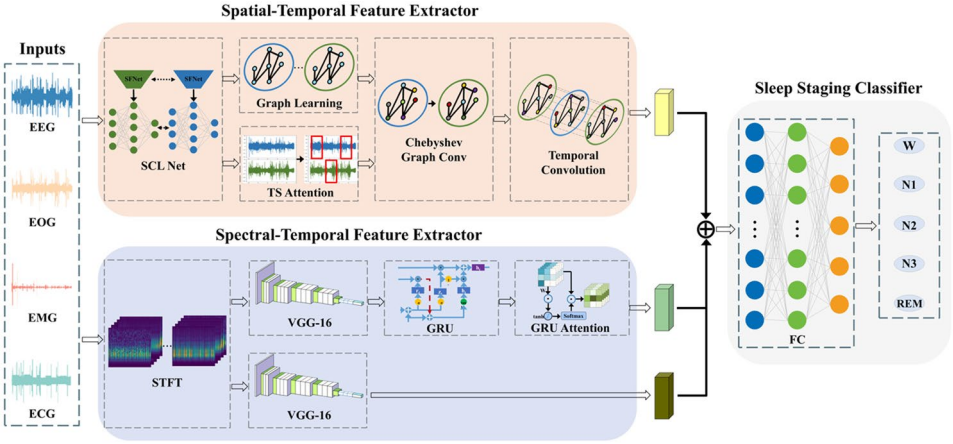

Fig. 1 The overall architecture of MSF-SleepNet. It includes three modules: a spatial–temporal feature extractor, spectral–temporal feature extractor, and sleep stage classifier

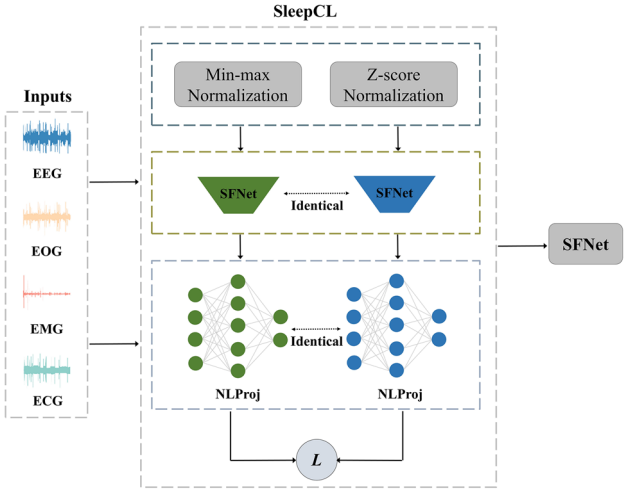

Fig. 2 The sleep contrastive learning network (SCL Net). In SleepCL, minimum-maximum normalization and Z-score normalization are jointly applied as data augmentation methods. SFNet is employed as the base encoder. The projection head comprises two fully connected layers and a ReLU activation function named NLProj. A cosine similarity calculation is selected to maximize similarity and minimize contrastive loss. SFNet is used for feature extraction

representation vectors $h_{\mathrm{i}}, h_{\mathrm{j}}$ from the augmented data sample. SFNet includes 1-D convolution layers, maxpooling layers, a ReLU activation function, batch normalization, and dropout. For the projection head, a single multi-layer perceptron (MLP) with a ReLU activation function was used for non-linear projection (NLProj). NLProj applies nonlinear transformation to project the representation h_{i} and h_{j} into representation Z_{i} and Z_{j}. For maximizing similarity and the contrastive loss function, a cosine similarity function is utilized to compute the similarity between two augmented signal data. The goal is to maximize the similarity between positive samples and minimize the similarity between negative samples. The formula for similarity is as follows:

$$\text{Similarity}(\boldsymbol{Z}_i, \boldsymbol{Z}_j) = \boldsymbol{Z}_i^{\top} \boldsymbol{Z}_j / \left(\|\boldsymbol{Z}_i\| \, \|\boldsymbol{Z}_j\|\right), \quad (1)$$

where $Z_i^{\mathrm{T}}, ||Z_i||, ||Z_j||$ are the transposes of Z_i, and the lengths of Z_i and Z_{ij}, respectively. The contrastive loss function and optimization process is illustrated below:

$$\begin{aligned} l(i,j) &= -\log \frac{\exp\left[Similarity(i,j)/\tau\right]}{\sum_{k=1 \wedge k \neq i}^{2N} \exp\left[Similarity(i,k)/\tau\right]} \\ \mathcal{L} &= \frac{1}{2N} \sum_{k=1}^{N} [l(2k-1, 2k) + l(2k, 2k-1)], \end{aligned} \quad (2)$$

where τ denotes the temperature hyperparameter, and N is the batch size of contrastive learning data. After the optimization process of the loss function, the distance between similar data in the projection space decreases while the distance between dissimilar data increases.

After SleepCL, SFNet is used to extract rich features of sleep stages. After SCL Net, the general features from unlabeled data are learned.

Spatial–temporal feature extraction

This module consists of graph structure learning, temporal attention and spatial attention, Chebyshev graph convolution, and temporal convolution. Each component is explained below.

Multi-modal physiological signals include signals from different physiological channels, such as EEG, EOG, EMG and ECG. These signals provide information about the brain, eyes, muscles, and heart, which together reflect an individual's overall physiological state and activity. When processing these multiple physiological signals, the graph structure data can be used to integrate and analyze the complex interaction between them. Therefore, our work adopts graph structure learning to construct and define the node and edge structures in the graph model according to the internal correlation between these signals, so as to represent the non-Euclidean space graph information represented by different modal physiological signals [19]. Graph structure learning is applied to learn and update the graph structure dynamically. The graph structure is defined as follows:

$$A_{ij}^{GL} = \frac{\exp\left(\mathrm{ReLU}\left(w^{\mathrm{T}} \left|x_i - x_j\right|\right)\right)}{\sum_{j=1}^{N} \exp\left(\mathrm{ReLU}\left(w^{\mathrm{T}} \left|x_i - x_j\right|\right)\right)} \quad (3)$$

where x_i, x_j represents any two nodes in the generated graph structure, ReLU is the non-linear activation function, and w is the weight matrix. The graph structure A_{ij}^{GL} is optimized by minimizing the following loss function:

$$L_{\mathrm{GL}} = \sum_{i,j=1}^{N} \left\|x_i - x_j\right\|_2^2 A_{ij}^{GL} + \alpha \left\|A^{GL}\right\|_F^2 \quad (4)$$

where α is the regularization parameter to control the degree of regularization. The absolute value in the formula represents the similarity of two nodes, and the greater the similarity, the higher the connection strength of two nodes in the graph structure. Temporal and spatial attention mechanisms [39] acquire the most relevant feature information of different sleep stages from temporal and spatial dimensions, respectively. Temporal attention and spatial attention are defined as follows:

$$T = V_t \cdot \sigma\left(\left(\boldsymbol{\mathcal{X}}^{(l-1)}\right)^{\mathrm{T}} M_1\right) M_2 \left(M_3 \boldsymbol{\mathcal{X}}^{(l-1)}\right) + b_t\Big), \quad (5)$$

$$S = V_s \cdot \sigma \left(\left(\boldsymbol{\mathcal{X}}^{(l-1)} \right) Z_1 \right) Z_2 \left(Z_3 \boldsymbol{\mathcal{X}}^{(l-1)} \right) \Big)^{\mathrm{T}} + b_s \Big), \quad (6)$$

where V_t, b_t, V_s, b_s, M_1, M_2, M_3, Z_1, Z_2, and Z_3 are learnable parameters; $\boldsymbol{\mathcal{X}}^{(l-1)}$ is the first l layer's input, and σ denotes the sigmoid activation function. In real-world scenarios, there are various types of graph data that describe complex relationships and interactions between entities in the form of graph structures. Traditional deep learning methods often struggle to process this data efficiently, as they are primarily designed for Euclidean data structures, such as images and text. However, this gap has been filled by the emergence of graph convolutional neural networks, a deep learning model specifically designed to process graph data. Compared to traditional convolutional neural networks (CNNs), GCNs can directly manipulate graph structures in non-Euclidean space. In our work, a Chebyshev graph convolution [18] is used in the GCN to process the graph data of time series, and the feature representation after convolution is output. The Chebyshev graph convolution can fully obtain the topological structure information of the graphs and realize the extraction of signal spatial features. The graph structure data generated by graph structure learning above is used as the input of Chebyshev graph convolution, and the number of nodes is determined by the number of nodes of the graph data, that is, the length of the signal fragment. The Chebyshev graph convolution is defined as follows:

$$\begin{aligned} g_\theta *_G x = g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})x \\ \tilde{L} = \frac{2}{\lambda_{\max}} L - \boldsymbol{I_N}, \end{aligned} \quad (7)$$

where g_θ, $*_G$, θ, x, T_k, $\lambda_{\max}$, and $\boldsymbol{I_N}$ denote the convolution kernel, graph convolution operation, a vector of Chebyshev coefficients, input data, Chebyshev polynomials recursively, a Laplacian matrix's maximum eigenvalue, and an identity matrix, respectively. And temporal convolution is employed to gain spatial–temporal features among neighboring sleep stages on the basis of spatial features. The temporal convolution is defined as follows:

$$\mathcal{Y}^{(l)} = \mathrm{ReLU} \left(\Phi * \left(\mathrm{ReLU} \left(g_\theta *_G \hat{\mathcal{Y}}^{(l-1)} \right) \right) \right), \quad (8)$$

where ReLU, Φ, and $*$ are the activation function, the parameters of the convolution kernel, and the convolution operation, respectively. Next, the spatial–temporal features are learned.

Spectral–temporal feature extractor

Spectral feature extraction

In order to learn spectral features, STFT and VGG-16 network are combined.

STFT is a basic time-frequency analysis method. STFT helps describe the frequency content of signal data at each point in time. The signal data is analyzed by STFT, and is then mapped to a two-dimensional function of frequency and time. To calculate STFT, the signal is divided into multiple short-time signal segments with some overlap by moving time window. The discrete Fourier Transform (DFT) is then applied for each segment to obtain each local spectrum. The STFT equation is

$$S(m,k) = \sum_{n=0}^{N-1} s\left(n + mN'\right) w(n) e^{-j\frac{2\pi}{N}nk}, \quad (9)$$

where $k = 0, 1, \ldots, N-1$, $S(m,k)$ denotes the m-index time–frequency spectrogram. N is window segment length. N' is the shifting step of the time window. $w(n)$ is the window of an N-point sequence.

In our network, we use STFT to transform the original time domain and frequency domain signals from raw data into time–frequency domain signals. Due to its good performance in extracting rich features, we apply VGG-16 [20] to the original signal data after signal processing in the feature extraction step to obtain spectral features from frequency information. After this step, the spectral features are learned.

Spectral–temporal feature extraction

Since there exists neighboring information of adjacent signal segments, we transform spectral features obtained by VGG-16 into five-timestep features to fully consider the potential correlation between neighboring data segments.

Then we apply a GRU [15] and GRU attention mechanism jointly to the five-timestep spectral features. This allows us to capture richer spectral–temporal feature representations from neighboring information of adjacent signal segments. The GRU is defined as follows:

$$h^{(t)} = \left(1 - z^{(t)}\right) \circ \tilde{h}^{(t)} + z^{(t)} \circ h^{(t-1)}, \quad (10)$$

where $z^{(t)}$ is the gating signal, $\tilde{h}^{(t)}$ is the information of the current signal, and $h^{(t-1)}$ is the information transmitted from the upper unit. The GRU attention mechanism allows us to attain the most relevant feature information among sleep physiological signals according to different weights. The GRU attention mechanism is defined as follows:

$$a_{g,t} = U_g \cdot \tanh\left(NX^{(l-1)} + b_g\right)$$
$$\boldsymbol{\alpha_{g,t}} = \frac{\exp\left(a_g\right)}{\sum_{k=1}^{T} \exp\left(a_k\right)}, \tag{11}$$

where U_g, N, and b_g are learnable parameters; $\tanh$ is the activation function; and $\alpha_{g,t}$ is the GRU attention matrix. After introducing GRU attention, the spectral–temporal features are learned.

Sleep stage classifier

After multi-stream feature extraction, we conduct a concatenate operation on three feature matrices to accomplish multi-stream feature fusion. Because a single segment without neighboring information can reflect the ontological characteristics of physiological signals, spectral features should be taken into consideration. The concatenate operation is defined as follows:

$$\mathcal{X} = \mathcal{X}^F \parallel \mathcal{X}^S \parallel \mathcal{X}^T, \tag{12}$$

where $\mathcal{X}^F$, $\mathcal{X}^S$, and $\mathcal{X}^T$ denote the spatial–temporal features, spectral features, and spectral–temporal features, respectively. $\parallel$ is the concatenate operation. Next, we feed the fused features into a 256-dimensional fully connected layer. After the Softmax activation layer, the output results are classified into five sleep stages. The pseudo-code of MSF-SleepNet is shown in Algorithm 1.

Algorithm 1 Pseudo-Code of MSF-SleepNet.

procedure MSF-SLEEPNET TRAINING
  Input: multi-modal physiological signal data
  Output: MSF-SleepNet model
  SFNet processes the input data to extract graph feature.
  SCL-Net enhances graph feature to obtain enhanced graph feature.
  Building graph data using enhanced feature and graph structure learning.
  Applying Chebyshev GCN and temporal CNN to graph data to obtain spatial-temporal feature.
  STFT transforms the input data to obtain spectrogram.
  VGG-16 extracts spectral feature from spectrogram.
  Using GRU and GRU attention to spectral feature to generate spectral-temporal feature.
  Combining spatial-temporal, spectral-temporal, and spectral features into fused features.
  Using classifier to predict sleep stages based on fused features.
  return MSF-SleepNet model
end procedure

Experiment results and discussion

Experiment settings

To fairly compare model performance, we utilize the same experimental setup and environment configuration for all models. On the ISRUC-S3 dataset, we employ a 10-fold cross-validation strategy, and use a subject-independent strategy for cross-validation. Specifically, we divide the data of 10 subjects into different training sets and testing sets, in which the data of each subject is used as the testing set, and the data of the remaining 9 subjects is used as the training set. Similarly, on the ISRUC-S1 dataset, the data of 100 subjects are randomly divided into ten groups, each group containing ten subjects' data, one of which is used as the testing set, and the data of the remaining nine groups is used as the training set. The MSF-SleepNet is implemented with Python 3.6, TensorFlow 1.15.0, and Keras 2.3.1. The experiments are performed on a computer server equipped with an Intel i7 CPU 3.40 GHz, 261 GB memory, Windows 10 operating system, and 4 × NVIDIA GeForce GTX TITAN X graphics cards with 12288 MiB GPU memory.

Datasets

Two PSG subsets of ISRUC datasets are employed in our experiments. The first one is ISRUC-S3, which contains 10 healthy subjects (9 males and 1 female) aged 30 to 58. Each PSG recording contains 12 channels: 6 EEG channels, 2 EOG channels, 3 EMG channels, and 1 ECG channel. The second subset is ISRUC-S1, which contains 100 subjects who suffer from sleep disorders (55 males and 45 females, aged 20 to 85). Each PSG recording contains the same 12 channels as in ISRUC-S3. The sleep stages corresponding to the PSG recordings are visually scored by two sleep disorder specialists. ISRUC-S3 and ISRUC-S1 are compared in Table 1.

Performance comparison

In this section, we compare our proposed network with other sleep staging methods (such as those employing an SVM, RF, CNN, RNN, or LSTM) for sleep stage classification on the datasets of ISRUC-S1 and ISRUC-S3 datasets. In this way, we can validate the performance of MSF-SleepNet. The evaluation metrics are overall accuracy, F1-score, and Kappa score,the F1-score of the evaluation metric is calculated separately for the five different sleep stages, including W, N1, N2, N3 and REM stages. Finally, the F1-score of these five classes are averaged, which is F1-score in the overall classification performance results. The results obtained by the different models are shown in Table 2. We compare the best experimental results of our model with the best performance of other methods. At the same time, we also conduct ten repeated experiments to calculate the mean value and standard deviation of our

Table 1 Comparison of datasets ISRUC-S1 and ISRUC-S3

Dataset	# of subjects	Gender		Age	Health condition	# of epochs					
		Male	Female			W	N1	N2	N3	REM	Total
ISRUC-S1	100	55	45	20–85	Sleep disorders	20098	11062	27511	17251	11265	87187
ISRUC-S3	10	9	1	30–58	Health	1651	1215	2609	2014	1060	8549

Table 2 Experimental results comparison of current methods on the ISRUC-S3 dataset

Paper	Models	Overall results			F1-score of different sleep stages				
		Accuracy	F1-score	Kappa	W	N1	N2	N3	REM
Alickovic et al. [9]	SVM	0.733	0.721	0.657	0.868	0.523	0.699	0.786	0.731
Memar et al. [25]	RF	0.729	0.708	0.648	0.858	0.473	0.704	0.809	0.699
Dong et al. [15]	MLP + LSTM	0.779	0.758	0.713	0.860	0.469	0.760	0.875	0.828
Supratak et al. [16]	CNN + BiLSTM	0.788	0.779	0.730	0.887	0.602	0.746	0.858	0.802
Chambon et al. [40]	CNN	0.781	0.768	0.720	0.870	0.550	0.760	0.851	0.809
Phan et al. [17]	ARNN + RNN	0.789	0.763	0.725	0.836	0.439	0.793	0.879	0.867
Jia et al. [18]	STGCN	0.799	0.787	0.741	0.878	0.574	0.776	0.864	0.841
Jia et al. [19]	MSTGCN	0.821	0.808	0.769	0.894	0.596	0.806	0.890	0.856
Li et al. [20]	MVF-SleepNet	0.841	0.828	0.795	0.900	0.625	0.833	**0.911**	0.873
Proposed model	MSF-SleepNet	**0.849**	**0.838**	**0.805**	**0.912**	**0.645**	**0.837**	0.910	**0.888**

Bold text denotes the best results for each evaluation indicator while underlined text denote the second best performance

network in each evaluation metric, and the results are as follows: the mean values of accuracy, F1-score and Kappa coefficient of the overall results are 0.846, 0.835 and 0.802, and the standard deviations are 0.0018, 0.0018 and 0.0024, respectively; the mean values of F1-score of the five different sleep stages are 0.909, 0.640, 0.836, 0.907 and 0.883, and the standard deviations are 0.0016, 0.0039, 0.0010, 0.0034 and 0.0041.

As can be seen from Table 2, in each evaluation metric, our model has better overall performance than other models, where the overall accuracy, F1-score and Kappa score are 0.849, 0.838 and 0.805, respectively. Traditional machine learning methods, such as those using an SVM or RF, fail to extract temporal and spatial features; meanwhile, deep learning models, like those using a CNN, RNN, or LSTM, are specialized in capturing relevant domain information of temporal and spatial dimensions. Hence, deep learning models achieve better classification performance on different sleep stages than traditional machine learning methods.

Although deep learning methods are better than most traditional methods, our proposed model outperforms these methods on almost all evaluation indicators. From the perspective of overall accuracy, F1-score and Kappa score, the classification performance are 1% better than the sub-optimal results. This is due to the fact that our method makes full use of and integrates multiple heterogeneous additional information such as unlabeled information, topological information, frequency information, and neighboring information; moreover, it fully considers spectral, temporal, and spatial features. For each specific sleep stage, our method is more accurate in classifying different sleep stages compared most other models. Specifically, the W stage and N3 stage have the highest classification accuracy 0.912 and 0.910 on account of the larger sample size and relatively obvious features compared to the other three sleep stages. In addition, N2 stage and REM stage are 0.4% and 1.5% higher than the sub-optimal model, 0.837 and 0.888, respectively. Our model's classification accuracy for the N1 stage is 2% higher than the second best classification result. However, the classification performance for the N1 stage is much lower than that for the other four sleep stages because it is the transitional phase between the W stage and N2 stage and its sample size is small due to fewer people sleeping the whole night. Overall, our model's experimental performance and classification accuracy is best, which validates its effectiveness and superiority for sleep stage classification.

From the classification results shown in Fig. 3, we can conclude that our proposed model can accurately classify most sleep stages to some extents. The confusion matrices show the performance results of MSF-SleepNet on two datasets, ISRUC-S1 and ISRUC-S3, which compare the true sleep stages with the predicted sleep stages. On the ISRUC-S3 dataset, MSF-SleepNet achieved more than 90% classification accuracy for the three sleep stages of W, N3 and REM, of which the classification accuracy of W stage reached 91.04%. At the same time, the classification performance of N2 stage reached more than 85%. On the ISRUC-S1 dataset, MSF-SleepNet has the highest classification accuracy of 90.26% among the five sleep stages, and the accuracy of N3 and REM stages also reached more than 85%. Therefore, the model can achieve excellent classification results on the whole. To research the influence of dataset size on classification performance, we conduct experiments on the ISRUC-S1 dataset with 50 subjects. The results are shown in Table 3. Table 3 demonstrates that our network has better classification performance compared to other models.

To discuss the classification performance of the proposed model on the PSG level and compare automatic classification with manual scoring by human experts, we visualize the whole night's PSG signal hypnogram of the subject that classified with the highest accuracy (91.93%) by our model. Figure 4 shows the true hypnogram (top) and predicted hypnogram (bottom) of these PSG data. As we can see, the W stage has the highest similarity

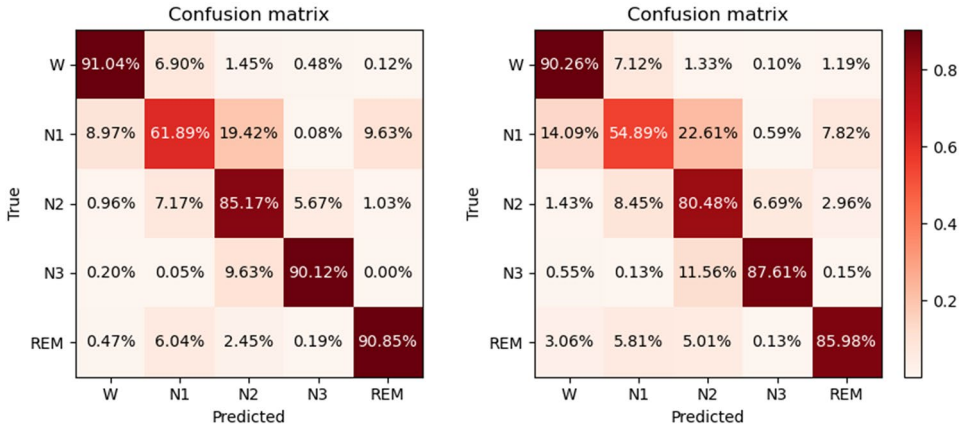

Fig. 3 Confusion matrices of the 10-fold cross-validation on ISRUC-S3 dataset (left) and ISRUC-S1 dataset (right)

Table 3 Experimental results comparison of current methods on ISRUC-S1 dataset

Paper	Models	Overall results			F1-score of different sleep stages				
		Accuracy	F1-score	Kappa	W	N1	N2	N3	REM
Alickovic et al. [9]	SVM	0.684	0.608	0.583	0.793	0.242	0.708	0808	0.490
Memar et al. [25]	RF	0.699	0.649	0.607	0.841	0.307	0.705	0.750	0.640
Dong et al. [15]	MLP + LSTM	0.703	0.648	0.614	0.807	0.301	0.724	0.817	0.591
Supratak et al [16].	CNN + BiLSTM	0.717	0.691	0.638	0.823	0.466	0.738	0.809	0.621
Supratak&YiKe [21]	CNN + RNN	0.778	0.758	0.714	0.883	0.532	0.764	0.848	0.763
Perslev et al. [22]	U-Net	0.770	0.770	–	0.890	0.520	0.790	0.770	**0.880**
Jia et al. [18]	STGCN	0.786	0.754	0.723	0.884	0.437	0.775	0.838	0.835
Jia et al. [19]	MSTGCN	0.804	0.785	0.748	0.887	0.545	0.791	<u>0.872</u>	0.832
Li et al. [20]	MVF-SleepNet	<u>0.821</u>	<u>0.802</u>	<u>0.768</u>	<u>0.908</u>	<u>0.562</u>	<u>0.811</u>	0.871	0.857
Proposed model	MSF-SleepNet	**0.826**	**0.809**	**0.774**	**0.912**	**0.570**	**0.812**	**0.884**	<u>0.865</u>

Bold text denotes the best results for each evaluation indicator while underlined text denotes the second best performance

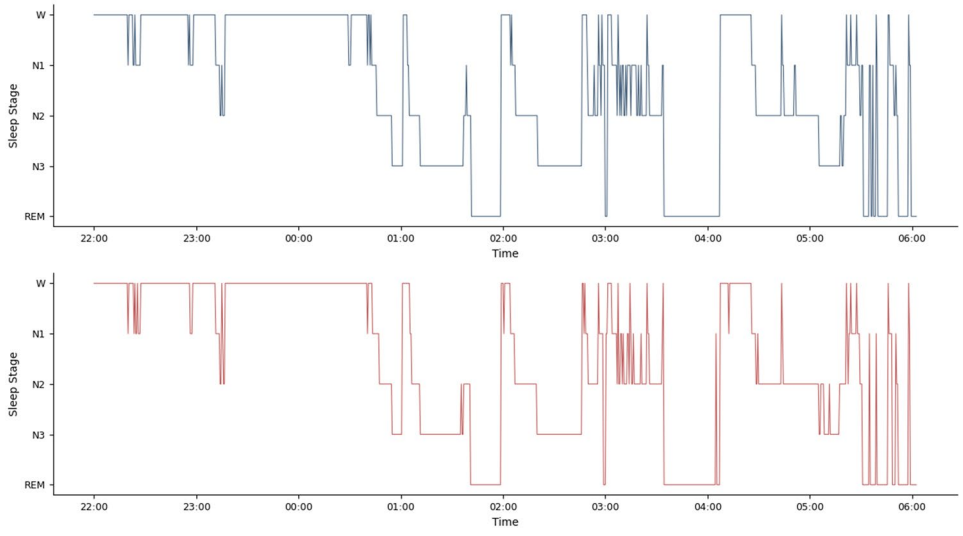

Fig. 4 Example hypnograms of ground truth (top) and prediction (bottom)

Table 4 Computational analysis of MSF-SleepNet

Parameter size	FLOPs	Training time
5.25 M	206.24M	54400s

Table 5 Comparison of FLOPs of MSF-SleepNet with other models

Models	FLOPs
AlexNet [41]	0.72G
SqueezeNet [42]	0.82G
SleepPrintNet [43]	48.31G
AttnSleepNet [44]	26.26G
SalientSleepNet [45]	226.92G
MMASleepNet [46]	18.23G
DynamicSleepNet [47]	12.46G
StAGN [48]	7.98G
Diff-SleepNet [49]	1.42G
cVAN [50]	0.47G
MSF-SleepNet	0.21G

between the true hypnogram and prediction hypnogram. Meanwhile, the similarity of the N1 stages is the lowest of all stages. Our method can correctly identify most stage transitions, such as W–N1, N2–N3, and N1–N2 transitions.

In order to analyze the computation complexity of our proposed model, we introduce two evaluation metrics: parameter size and floating-point operations (FLOPs). The parameter size reflects the total number of model parameters and serves as an indicator of the model's suitability for deployment on storage-constrained edge devices. FLOPs quantify the computational workload required for MSF-SleepNet to perform classification tasks. Each of these metrics provides valuable insights into different aspects of the model's performance and resource requirements. Detailed results are shown in Table 4. From the data in the table, parameter size, FLOPs and training time of MSF-SleepNet are 5.25 M, 206.24 M and 54400s respectively. Taken together, the lower parameter size, FLOPs, and training time make the model MSF-SleepNet more suitable for deployment in resource-limited environments, improving the utility and availability of the model, while also leading to higher efficiency and lower energy consumption. At the same time, in order to show the advantages of our proposed model in low computational complexity, we also compared the FLOPs metric in ISRUC dataset with other sleep stage classification models, and the comparison results are shown in the Table 5. In order to facilitate the comparison, we unified the unit of FLOPs. According to the results in the table, compared with the current cutting-edge sleep stage classification methods, our proposed model achieves the lowest FLOPs, which further indicates the superiority of MSF-SleepNet in computational complexity. It is proved that the model achieves a good balance between sleep stage classification performance and computational complexity.

Model interpretability

Although deep learning models have good performance, one of their main weaknesses is that they are black-box models, which makes them unusable in clinical settings. When it comes to the application of artificial intelligence in healthcare and medicine, people generally pay attention to the potential of deep learning algorithms. Understanding the underlying mechanisms of these models and improving their interpretability can help sleep specialists make more reliable and confident decisions.

In our study, Local Interpretable Model-Agnostic Explanations (LIME) [51] is used as an interpretability tool to explain the outputs of our proposed model. The LIME is a post-hoc interpretability tool. LIME explains the behavior of the black-box model based on a linear model around the instance of interest [52]. LIME needs data instances and trained model as input. The signal input is perturbed, and the trained model produces a prediction of the data instance. LIME determines the attributes and their values that play an important role in generating a particular prediction.

Experiments are performed to verify the predictions of our model for each sleep stage. Two examples of LIME outputs are displayed in Fig. 5. In our experiments, we set the number of features as 10. The y-axis shows the 10 features that are predicted to have the greatest impact on the corresponding class and their corresponding value ranges. The x-axis displays the predicted weight value of each feature on the corresponding class. The right bars represent important features that make the model prediction for this class, while the left ones represent features predicted for other classes; Bar lengths represent the weight given to each prediction. Figure 5 demonstrates the true predictions of our model and their corresponding significant features.

In Fig. 6, we visualize the input data and the 10 important features of their corresponding LIME outputs. The input data is PSG signals after a short-time Fourier transform.

There are five main brain waves of EEG signals differentiated by different frequency bands; from low to high frequency, these bands are δ (delta, 0.16–3.99 Hz), θ (theta, 4–7.99 Hz), α (alpha, 8–11.99 Hz), σ (sigma, 12–15.99 Hz), and β (beta, 16–30 Hz) [25]. According to the AASM manual, these five frequency bands are highly indicative for each specific sleep stage. Table 6 displays this indicative relevance in the frequency domain. In order to determine whether the predictions of our model conform to this domain knowledge, we perform frequency domain occlusion [53] on EEG signals and investigate the change in prediction results during the

Fig. 5 LIME outputs of two instances predicted as N2 and REM

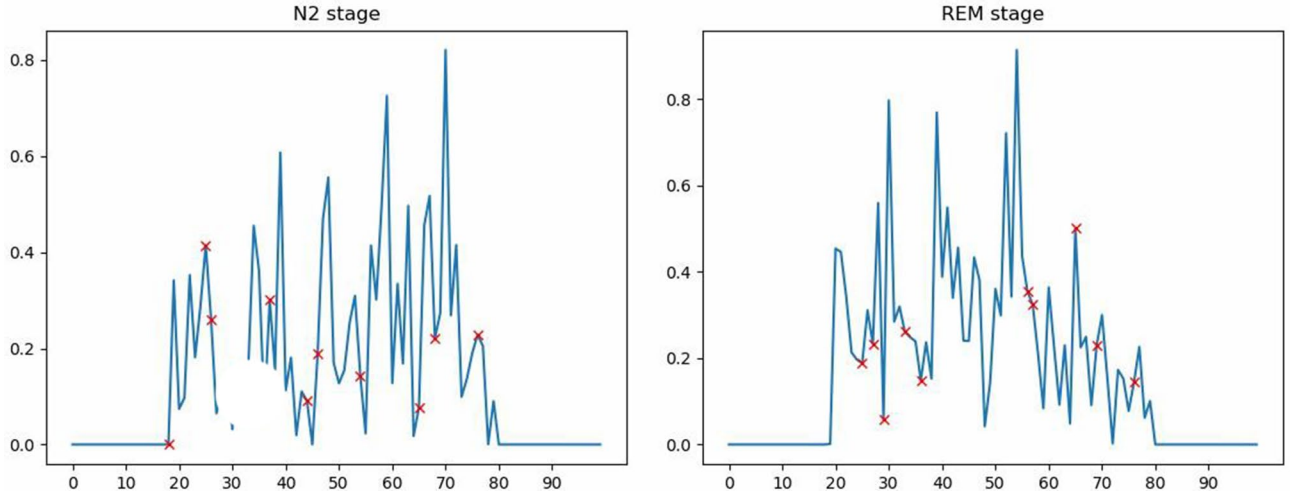

Fig. 6 Input data shows a N2 stage and an REM stage

Table 6 The characteristic frequency bands of different sleep stages in frequency domain

Sleep stages	W	N1	N2	N3	REM
Frequency domain	α, σ, β	θ, α	θ, σ	δ	θ, α

occlusion process. According to the domain knowledge, omitting some frequency bands of EEG signals may result in a decrease in prediction accuracy for certain sleep stages. We occlude one or several frequency bands in our experiments at a time. Concretely, a band-stop filter is used to delete a specific frequency in a given range.

Figure 7 shows the confusion matrix results of our frequency domain occlusion experiments. When removing δ frequencies, the prediction accuracy for N3 is significantly decreased. This shows that δ is the characteristic frequency band of N3, which conforms to the domain knowledge presented in Table 6. While omitting θ frequencies in occlusion experiments, it is clear that the accurate predictions of N1, N2, and REM reduce more than for the other two sleep stages. This shows that θ is the characteristic frequency band of N1, N2, and REM, which is consistent with the AASM manual. Furthermore, we occlude several frequency bands at a time to investigate the effect on predictions for each sleep stage. If occluding α, β, and σ frequencies, the prediction accuracies of W and N2 decrease. This shows that α, β, and σ jointly influence the prediction of W and N2 stages, which conforms with the AASM manual. When we omit θ and α frequencies, the TPR of W, N1, and N2 achieves decreases by 2–3%. Therefore, θ and α play important roles in predicting these sleep stages; this result is in accordance with that presented in Table 6. A similar conclusion can be found when we remove θ and σ frequencies: the TPR of N2 obviously decreases more than that of the other stages. This shows that θ and σ are the characteristic frequency bands of N2; this finding also aligns with the AASM manual.

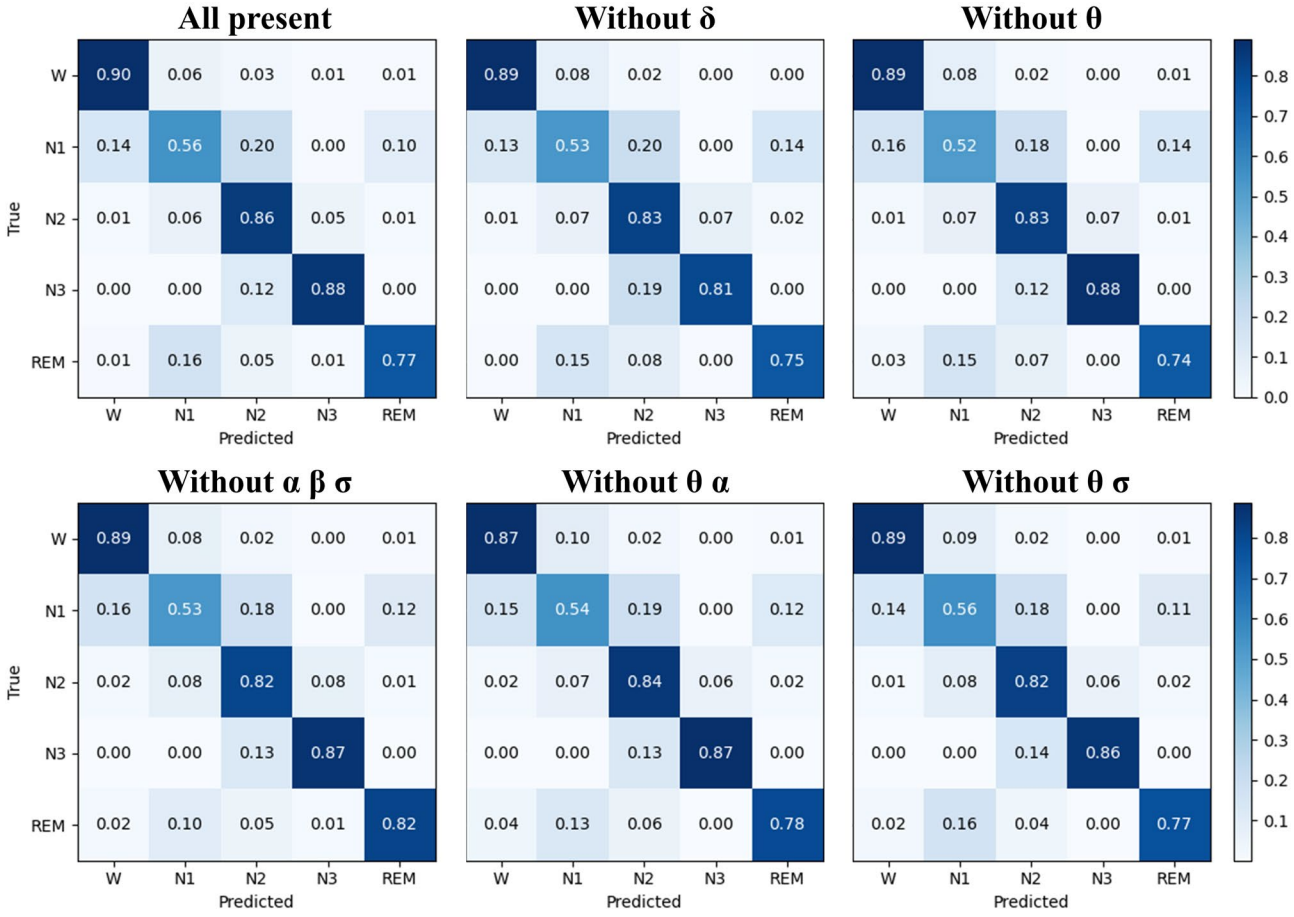

Fig. 7 Frequency domain occlusion. Confusion matrices with five frequency bands all present (upper left) or occluding one or several concrete bands (all other images)

Ablation experiments

To validate the effect of each module in our network, we decompose our interpretable multi-stream fusion network into five independent parts. Then, we design ablation experiments to verify the validity of each part. Each variant module is constructed as described below.

- Module A (SFNet): 1-D convolution neural network named SFNet is employed as the base model in our ablation experiments.
- Module B (+SleepCL): A contrastive learning method is applied to the SFNet into SCLNet.
- Module C (+Graph Learning, Attention Mechanism, Chebyshev Graph and Temporal Convolution): On the basis of SCLNet, a temporal and spatial attention mechanism, graph structure learning, Chebyshev graph convolution neural network, and temporal convolution are added to construct the first stream of our fusion model.
- Module D (+STFT and VGG-16): Based on module C, the short-time Fourier transform and VGG-16 are appended to fuse the first and second streams of the model.
- Module E (+Five-Timestep Features, GRU, and GRU Attention Mechanism): On the basis of module D, five-timestep features are fed to a GRU and GRU attention mechanism to build the whole of our model.

These five modules play key roles in our proposed model, as illustrated in Figs. 8 and 9. As seen in Fig. 8, with the continuous fusion of modules, the three evaluation metrics gradually increase. From modules A to E, the overall accuracy is 0.784, 0.813, 0.821, 0.831, and 0.849, and the F1-scores are 0.762, 0.796, 0.807, 0.818, and 0.838, respectively. Lastly, Kappa scores are 0.723, 0.759, 0.769, 0.782, and 0.805, respectively. To sum up, these results demonstrate the validity of each module. The same conclusion is illustrated in Fig. 9, where we see an increase in F1-score for each class' classification from modules A to E. For example, the F1-scores for classifying the N3 stage are 0.866, 0.886, 0.888, 0.899, and 0.910, respectively. The same rising trend was observed for the other four sleep

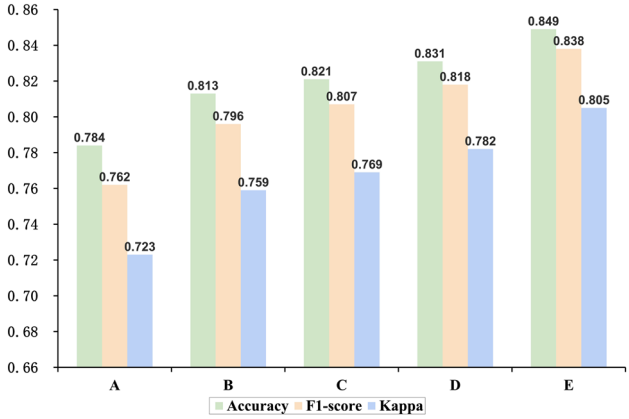

Fig. 8 Comparison of each module to validate their effect on the whole model

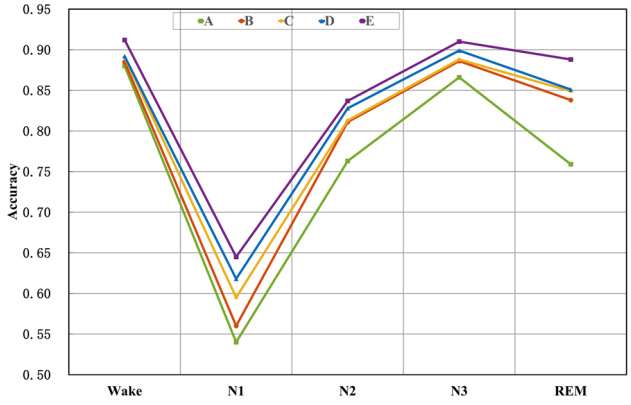

Fig. 9 Comparison of each module in terms of F1-score for classification of five classes

stages. This further shows the effectiveness of each module in our network.

Conclusion

In this study, we propose an interpretable multi-stream fusion network with contrastive learning for sleep stage classification. This network is named MSF-SleepNet, which is composed of three parts: a spatial–temporal feature extractor, a spectral–temporal feature extractor, and a sleep stage classifier. The spatial–temporal feature extractor is composed of Chebyshev graph convolution and temporal convolution to obtain spatial–temporal features from body-topological information. For the spectral–temporal feature extractor, STFT and a GRU are utilized to learn the spectral–temporal features. By fusing the spatial–temporal and spectral–temporal features, the sleep stage classifier uses a contrastive learning scheme to enhance differences in feature patterns of sleep monitoring signals across various sleep stages. LIME is employed to improve MSF-SleepNet's interpretability. The experimental results show that MSF-SleepNet outperforms cutting-edge methods: it has an accuracy of 0.849 and F1-score of 0.838 on the ISRUC-S3 dataset, and an accuracy of 0.826 and F1-score of 0.809 on the ISRUC-S1 dataset. Our study demonstrates that fusing multiple features from heterogeneous additional information can significantly improve sleep stage classification performance. In addition, the interpretability analysis of MSF-SleepNet further illustrates its reliability and transparency; moreover, the model's predicted output is in accordance with the AASM manual and domain knowledge.

Acknowledgements

We note that a shorter conference version of this paper appeared in the 2022 IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM 2022). Compared with the preliminary presented version, this manuscript supplements additional large number of experiments, effectiveness validation of model components, and interpretability analysis. It includes but not limits to: one more dataset of ISRUC-S1 is utilized to further validated the sleep staging performance of MSF-SleepNet; ablation study is conducted to verify the effectiveness of each parts in MSF-SleepNet; The LIME is utilized to conduct the interpretability of MSF-SleepNet.

Author contributions

XMF and WJM conceived and designed the study. JX, RQG, and RXW prepared the experimental equipment and resources. JRC analyzed the data. XMF, WJM, and YL interpreted the results. JRC wrote the manuscript. All authors read and approved the manuscript.

Funding

This work is partially supported by the National Natural Science Foundation of China under grant No. 62473267, the Guangdong Basic and Applied Basic Research Foundation under grant No. 2022B1515130009 and 2025A1515011614, the Natural Science Foundation of Top Talent of SZTU under grant No.GDRC202318, and the Guangzhou Municipal Key Research and Development Program Fund under grant No. 2025B03J0019 and 2023B03J0172.

Data availability

The data supporting the findings of this study comes from publicly datasets that are available for use. At the same time, upon reasonable request and permission to use.

Declarations

Ethics approval and consent to participate

The datasets used in study can be publicly available at https://sleeptight.isr.uc.pt/.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 14 March 2024 / Accepted: 4 April 2025
Published online: 14 April 2025

References

1. Kim CE, Shin S, Lee H-W, Lim J, Lee J-K, Shin A, Kang D. Association between sleep duration and metabolic syndrome: a cross-sectional study. BMC Public Health. 2018;18(1):1–8.
2. Kong G, Li C, Peng H, Han Z, Qiao H. EEG-based sleep stage classification via neural architecture search. IEEE Trans. Neural Syst. Rehabil. Eng. 2023;31:1075–85.
3. Dorokhov VB, Taranov AO, Sakharov DS, Gruzdeva SS, Tkachenko ON, Sveshnikov DS, Bakaeva ZB, Putilov AA. Linking stages of non-rapid eye movement

sleep to the spectral EEG markers of the drives for sleep and wake. J. Neurophysiol. 2021;126(6):1991–2000.
4. Boostani R, Karimzadeh F, Nami M. A comparative review on sleep stage classification methods in patients and healthy individuals. Comput. Methods Programs Biomed. 2017;140:77–91.
5. Li R, Wang B, Zhang T, Sugi T. A developed LSTM-ladder-network-based model for sleep stage classification. IEEE Trans. Neural Syst. Rehabil. Eng. 2023;31:1418–28.
6. Wassan JT, Wang H, Browne F, Zheng H. A comprehensive study on predicting functional role of metagenomes using machine learning methods. IEEE/ACM Trans. Comput. Biol. Bioinf. 2018;16(3):751–63.
7. Chen Z, Yang Z, Wang D, Huang M, Ono N, Altaf-Ul-Amin M, Kanaya S. An end-to-end sleep staging simulator based on mixed deep neural networks. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2021. p. 848–53.
8. Taran S, Sharma PC, Bajaj V. Automatic sleep stages classification using optimize flexible analytic wavelet transform. Knowledge-Based Syst. 2020;192. https://doi.org/10.1016/j.knosys.2019.105367.
9. Alickovic E, Subasi A. Ensemble SVM method for automatic sleep stage classification. IEEE Trans. Instrum. Meas. 2018;67(6):1258–65.
10. Klok AB, Edin J, Cesari M, Olesen AN, Jennum P, Sorensen HB. A new fully automated random-forest algorithm for sleep staging. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE; 2018. p. 4920–23.
11. Smith A, Anand H, Milosavljevic S, Rentschler KM, Pocivavsek A, Valafar H. Application of Machine Learning to sleep stage classification. In: 2021 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE; 2021. p. 349–54.
12. Dimitriadis SI, Salis C, Linden D.A novel, fast and efficient single-sensor automatic sleep-stage classification based on complementary cross-frequency coupling estimates. Clin. Neurophysiol. 2018;129(4):815–28.
13. Shahbakhti M, Beiramvand M, Eigirdas T, Solé-Casals J, Wierzchon M, Broniec-Wojcik A, Augustyniak P, Marozas V. Discrimination of wakefulness from sleep stage I using nonlinear features of a single frontal EEG channel. IEEE Sens. J. 2022;22(7):6975–84.
14. Zhang T, Jiang Z, Li D, Wei X, Guo B, Huang W, Xu G. Sleep staging using plausibility score: a novel feature selection method based on metric learning. IEEE J. Biomed. Health Inf. 2020;25(2):577–90.
15. Dong H, Supratak A, Pan W, Wu C, Matthews PM, Guo Y. Mixed neural network approach for temporal sleep stage classification. IEEE Trans. Neural Syst. Rehabil. Eng. 2017;26(2):324–33.
16. Supratak A, Dong H, Wu C, Guo Y. DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG. IEEE Trans. Neural Syst. Rehabil. Eng. 2017;25(11):1998–2008.
17. Phan H, Andreotti F, Cooray N, Chén OY, De Vos M. SeqSleepNet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. IEEE Trans. Neural Syst. Rehabil. Eng. 2019;27(3):400–10.
18. Jia Z, Lin Y, Wang J, Zhou R, Ning X, He Y, Zhao Y. GraphSleepNet: adaptive spatial-temporal graph convolutional networks for sleep stage classification. Ijcai. 2020;1324–30.
19. Jia Z, Lin Y, Wang J, Ning X, He Y, Zhou R, Zhou Y, Li-wei HL. Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification. IEEE Trans. Neural Syst. Rehabil. Eng. 2021;29:1977–86.
20. Li Y, Chen J, Ma W, Zhao G, Fan X. MVF-SleepNet: multi-view fusion network for sleep stage classification. IEEE J. Biomed. Health Inf. 2022. doi:https://doi.org/10.1109/JBHI.2022.3208314.
21. Supratak A, Guo Y. TinySleepNet: an efficient deep learning model for sleep stage scoring based on raw single-channel EEG. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2020. p. 641–44.
22. Perslev M, Darkner S, Kempfner L, Nikolic M, Jennum PJ, Igel C. U-Sleep: resilient high-frequency sleep staging. Npj Digital Med. 2021;4(1):1–12.
23. Khalighi S, Sousa T, Santos JM, Nunes U. ISRUC-Sleep: a comprehensive public dataset for sleep researchers. Comput. Methods Programs Biomed. 2016;124:180–92.
24. Sekkal RN, Bereksi-Reguig F, Ruiz-Fernandez D, Dib N, Sekkal S. Automatic sleep stage classification: from classical machine learning methods to deep learning. Biomed. Signal Process. Control. 2022;77. https://doi.org/10.1016/j.bspc.2022.103751.
25. Memar P, Faradji F.A novel multi-class EEG-based sleep stage classification system. IEEE Trans. Neural Syst. Rehabil. Eng. 2017;26(1):84–95.
26. Dhok S, Pimpalkhute V, Chandurkar A, Bhurane AA, Sharma M, Acharya UR. Automated phase classification in cyclic alternating patterns in sleep stages using Wigner–Ville distribution based features. Comput. Biol. Med. 2020;119. https://doi.org/10.1016/j.compbiomed.2020.103691.
27. Gunnarsdottir KM, Gamaldo CE, Salas RM, Ewen JB, Allen RP, Sarma SV. A novel sleep stage scoring system: combining expert-based rules with a decision tree classifier. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2018. p. 3240–43.
28. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. Neurocomputing. 2018;300:70–79.
29. Zhang J, Wu Y. Competition convolutional neural network for sleep stage classification. Biomed. Signal Process. Control. 2021;64. https://doi.org/10.1016/j.bspc.2020.102318.
30. Yang Y, Su X, Zhao B, Li G, Hu P, Zhang J, Hu L. Fuzzy-based deep attributed graph clustering. IEEE Trans. Fuzzy Syst. 2023;32(4):1951–64.
31. Kwon HB, Choi SH, Lee D, Son D, Yoon H, Lee MH, Lee YJ, Park KS. Attention-based LSTM for non-contact sleep stage classification using IR-UWB radar. IEEE J. Biomed. Health Inf. 2021;25(10):3844–53.
32. Sokolovsky M, Guerrero F, Paisarnsrisomsuk S, Ruiz C, Alvarez SA. Deep learning for automated feature discovery and classification of sleep stages. IEEE/ACM Trans. Comput. Biol. Bioinf. 2019;17(6):1835–45.
33. Phan H, Andreotti F, Cooray N, Chén OY, De Vos M. Joint classification and prediction CNN framework for automatic sleep stage classification. IEEE Trans. Biomed. Eng. 2018;66(5):1285–96.
34. Li Y, Luo S, Zhang H, Zhang Y, Zhang Y, Lo B. MtCLSS: multi-task contrastive learning for semi-supervised pediatric sleep staging. IEEE J. Biomed. Health Inf. 2022;27(6):2647–55. https://doi.org/10.1109/JBHI.2022.3213171.
35. Liu Y, Wu J, Qu L, Gan T, Yin J, Nie L. Self-supervised correlation learning for cross-modal retrieval. IEEE Trans. Multimedia. 2022;25:2851–63. https://doi.org/10.1109/TMM.2022.3152086.
36. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: Proceedings of the 37th International Conference on Machine Learning. PMLR; 2020. Vol. 119, p. 1597–607,.
37. Chen J, Li Y, Xiao J, Ge R, Ma W, Fan X. MSF-SleepNet: multi-stream fusion network with contrastive learning for sleep stage classification. In: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2022. p. 1493–96.
38. Ellis CA, Zhang R, Carbajal DA, Miller RL, Calhoun VD, Wang MD. Explainable sleep stage classification with multimodal electrophysiology time-series. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2021. p. 2363–66.
39. Fu Y, Wang X, Wei Y, Huang T. Sta: spatial-temporal attention for large-scale video-based person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2019. vol. 33, p. 8287–94.
40. Chambon S, Galtier MN, Arnal PJ, Wainrib G, Gramfort A. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. IEEE Trans. Neural Syst. Rehabil. Eng. 2018;26(4):758–69.
41. Ismail Fawaz H, Lucas B, Forestier G, Pelletier C, Schmidt DF, Weber J, Webb GI, Idoumghar L, Muller P-A, Petitjean F. Inceptiontime: finding alexnet for time series classification. Data Min. Knowl. Discov. 2020;34(6):1936–62.
42. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. Squeezenet: alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. arXiv preprint arXiv:1602.07360. 2016.
43. Jia Z, Cai X, Zheng G, Wang J, Lin Y. Sleepprintnet: a multivariate multimodal neural network based on physiological time-series for automatic sleep staging. IEEE Trans. Artif. Intell. 2020;1(3):248–57.
44. 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 Trans. Neural Syst. Rehabil. Eng. 2021;29:809–18.
45. Jia Z, Lin Y, Wang J, Wang X, Xie P, Zhang Y. Salientsleepnet: multimodal salient wave detection network for sleep staging. arXiv preprint arXiv:2105.13864. 2021.
46. Yubo Z, Yingying L, Bing Z, Lin Z, Lei L. Mmasleepnet: a multimodal attention network based on electrophysiological signals for automatic sleep staging. Front. Neurosci. 2022;16:973761.
47. Wenjian W, Qian X, Jun X, Zhikun H. Dynamicsleepnet: a multi-exit neural network with adaptive inference time for sleep stage classification. Front. Physiol. 2023;14:1171467.
48. Chen J, Dai Y, Chen X, Shen Y, Luximon Y, Wang H, He Y, Ma W, Fan X. Stagn: spatial-temporal adaptive graph network via contrastive learning for sleep

stage classification. In Proceedings of the 2023 SIAM International Conference on Data Mining (SDM). SIAM; 2023. p. 199–207.
49. Xu X, Cong F, Chen Y, Chen J. Sleep stage classification with multi-modal fusion and denoising diffusion model. IEEE J. Biomed. Health Inf. 2024.
50. Yang Z, Qiu M, Fan X, Dai G, Ma W, Peng X, Fu X, Li Y. cvan: a novel sleep staging method via cross-view alignment network. IEEE J. Biomed. Health Inf. 2024.
51. Lee W, Kim G, Yu J, Kim Y. Model interpretation considering both time and frequency axes given time series data. Appl Sci. 2022;12(24). https://doi.org/10.3390/app122412807.
52. Troncoso-Garca A, Martnez-Ballesteros M, Martnez-Álvarez F, Troncoso A. Explainable machine learning for sleep apnea prediction. Procedia Comput. Sci. 2022;207:2930–39.
53. Pathak S, Lu C, Nagaraj SB, van Putten M, Seifert C. STQS: interpretable multi-modal spatial-temporal-sequential model for automatic sleep scoring. Artif. Intell. Med. 2021;114:102038. https://doi.org/10.1016/j.artmed.2021.102038.

Publisher's Note