Multi-View Spatial-Temporal Graph Convolutional Networks With Domain Generalization for Sleep Stage Classification

Ziyu Jia,

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Youfang Lin,

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Jing Wang [Member, IEEE],

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Xiaojun Ning,

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Yuanlai He,

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Ronghao Zhou,

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Yuhan Zhou,

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Li-wei H. Lehman

Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA

Abstract

Sleep stage classification is essential for sleep assessment and disease diagnosis. Although previous attempts to classify sleep stages have achieved high classification performance, several challenges remain open: 1) How to effectively utilize time-varying spatial and temporal features from multi-channel brain signals remains challenging. Prior works have not been able to fully

Corresponding author: Jing Wang., wj@bjtu.edu.cn.

utilize the spatial topological information among brain regions. 2) Due to the many differences found in individual biological signals, how to overcome the differences of subjects and improve the generalization of deep neural networks is important. 3) Most deep learning methods ignore the interpretability of the model to the brain. To address the above challenges, we propose a multi-view spatial-temporal graph convolutional networks (MSTGCN) with domain generalization for sleep stage classification. Specifically, we construct two brain view graphs for MSTGCN based on the functional connectivity and physical distance proximity of the brain regions. The MSTGCN consists of graph convolutions for extracting spatial features and temporal convolutions for capturing the transition rules among sleep stages. In addition, attention mechanism is employed for capturing the most relevant spatial-temporal information for sleep stage classification. Finally, domain generalization and MSTGCN are integrated into a unified framework to extract subject-invariant sleep features. Experiments on two public datasets demonstrate that the proposed model outperforms the state-of-the-art baselines.

Keywords

Sleep stage classification; spatial-temporal graph convolution; transfer learning; domain generalization

I.   INTRODUCTION

SLEEP stage classification is important for the assessment of sleep quality and the diagnosis of sleep disorders. Sleep experts identify sleep stages based on American Academy of Sleep Medicine (AASM) standard [1] and observations recorded in polysomnography (PSG), which includes electroencephalography (EEG) at different positions on the head and electrooculography (EOG). The transition rules among different sleep stages recorded in the AASM standard, which can assist sleep experts in identifying the sleep stages. Although these rules provide valuable information, classifying the sleep stages by human sleep experts is still a tedious and time-consuming task. Moreover, the classification results are affected by the variability and subjectivity of sleep experts.

Automatic sleep stage classification can greatly improve the efficiency of traditional sleep stage classification and has important clinical value. Many researchers have made great contributions to automate this classification task. At first, traditional machine learning methods based on time domain, frequency domain, and time-frequency domain features are adopted [2], [3]. However, the classification accuracy of these methods depends heavily on feature engineering and feature selection, which require substantial expert knowledge. Recently, deep learning methods have been widely applied to automatically classify sleep stage thanks to its powerful ability of representation learning. For example, Convolutional Neural Network (CNN) [4] and Recurrent Neural Network (RNN) [5] are often utilized to learn appropriate feature representations from transformed data or directly from raw data.

Although the existing methods [6]–[10] achieve high accuracy for sleep stage classification, these methods have not sufficiently solved the following challenges: 1) The spatial-temporal features of sleep stages have not been fully considered. In particular, the topology among brain regions has not been effectively employed to capture richer spatial features.

2) Physiological signals vary significantly across different subjects, which hinders the generalizability of the trained classifiers. 3) Most deep learning methods, especially related graph neural network models, ignore the importance of model interpretability to the brain.

There have been several attempts to address *the first challenge* [7], [11]–[13]. For example, CNN is usually applied to extract the spatial features of the brain, and RNN is applied to capture temporal features during sleep transition. However, the limitation of these networks is that their input must be grid data (image-like representations) without utilizing the connections among brain regions [14]. Due to the fact that brain regions are in non-Euclidean space, graph is the most appropriate data structure to indicate brain connection. Therefore, GraphSleepNet [15] is proposed to classify sleep stages based on the functional connectivity of the brain network and using spatial-temporal graph convolution to achieve the state-of-the-art performance. However, in the brain network based on functional connectivity, there may not necessarily be connections among physically adjacent brain regions. In fact, existing neuroscience research shows that brain regions that are adjacent to each other at physical distances can influence each other [16]. However, GraphSleepNet only utilizes the functional connectivity of the brain to construct the sleep stage networks, which ignores the importance of the physical proximity of the brain in space. For *the second challenge*, some researchers try to apply transfer learning methods to improve the generalization of the models [17], [18]. The existing sleep stage classification models based on transfer learning are all two-step training paradigms. That is, these models need to be pre-trained and then fine-tuned to new subject data. The fine-tuning operation needs to collect sleep data from specific new subjects or datasets, which is quite expensive and inconvenient. In addition, the generalization of transfer learning models that need to be fine-tuned is limited. These models are designed for specific subjects and may not show excellent performance on other new subjects. Therefore, fine-tuning is only applicable to the personalized (subject-variant) model of the specific subject. And whenever a new subject needs to be evaluated, the existing model must be re-collected and re-trained. Therefore, for clinical systems suitable for unknown users, fine-tuning may become inefficient. For *the third challenge*, previous attempts to develop interpretable CNN or RNN classification models have been sparse [7], [12], [19]. Specifically, no attempt has been made to interpret the key modules of graph neural network for sleep stage classification from the perspective of the brain network.

In order to address the above challenges, we propose the multi-view spatial-temporal graph convolutional networks (MSTGCN) with domain generalization for sleep stage classification. Figure 1 illustrates the overall architecture of our model. Specifically, 1) we construct two brain view graphs based on the spatial proximity and functional connectivity of the brain, where each EEG channel corresponds to a node of the graph, and the specific connections among the channels correspond to the edge of the graph. 2) Then, we utilize spatial graph convolution to capture rich spatial features. Temporal convolution is applied for capturing the transition among different sleep stages. Actually, sleep experts usually identify the class label of one sleep state according to both the characteristic EEG waves of the current state and the class labels of its neighbors. 3) We design a spatial-temporal attention mechanism to capture the most relevant spatial-temporal information on the sleep stages. 4) Finally, we apply the adversarial domain generalization, which is a typical method

of transfer learning without fine-tuning. In the process of model training, each subject is employed as a specific source domain for subject-invariant sleep feature extraction. The subject-invariant sleep feature does not vary with different subjects and is related to sleep stage classification. The advantage of the domain generalization is that it does not require any information in the new subjects (target domain).

To the best of our knowledge, it is the first attempt to apply spatial-temporal graph neural networks with domain generalization for sleep stage classification. Overall, the main contributions of the proposed model for sleep stage classification are summarized as follows:

- We construct different brain views based on the functional connectivity and physical distance proximity of the brain. The complementarity of different views provides rich spatial topology information for classification tasks.

- We design a spatial-temporal graph convolution with attention mechanism, which consists of spatial-temporal graph convolution for spatial-temporal features and attention mechanism for capturing the most relevant spatial-temporal information for sleep stage classification.

- We integrate domain generalization and spatial-temporal graph convolutional networks into a unified framework to extract subject-invariant sleep features.

- We conduct experiments on two public sleep datasets, namely ISRUC-S3 and MASS-SS3. Experimental results demonstrate that the proposed model achieves the state-of-the-art performance.

- We explore the interpretability of the key modules of the model. In particular, we present the functional connectivity obtained through adaptive graph learning. The results indicate that functional connectivity during light sleep is more complex than that during deep sleep.

Compared to the Adaptive Spatial-Temporal Graph Convolutional Networks (called GraphSleepNet) published in our preliminary work [15], MSTGCN has the following important improvements: 1) The brain network based on physical distance proximity is constructed. It and the preliminary adaptive functional connectivity brain network form a multi-view brain network, which can provide rich brain spatial topology information for sleep stage classification. 2) Domain generalization is integrated with spatial-temporal graph convolutional networks into a unified framework to improve the generalization of the proposed model. 3) Experiments are conducted to evaluate the effectiveness of MSTGCN on two sleep datasets, of which ISRUC-S3 is not evaluated in our preliminary work. Moreover, we conduct the ablation experiments to evaluate the impact of each component of MSTGCN on the performance. 4) The interpretability of the key modules in MSTGCN is explored and discussed.

II. RELATED WORK

In recent years, time series analysis has attracted the attention of many researchers [20], [21]. As a typical time series, physiological signals are used in many fields, such as motor

imagery [22]–[24], emotion recognition [25], [26], and sleep stage classification [15], etc. With the development of deep learning, two popular deep learning models, CNN and RNN, are widely applied in sleep stage classification. Specifically, a fast discriminative complex-valued CNN (FDCCNN) [27] is proposed to capture the sleep information hidden inside EEG signals. A CNN model based on multivariate and multimodal physiological signals [7] takes into account the transitional rules of sleep stages to assist classification. A hierarchical RNN named SeqSleepNet [13] tackles the task as a sequence-to-sequence classification task. At the same time, hybrid models are also employed by some researchers. DeepSleepNet [12] utilizes CNN to extract time-invariant features, and Bi-directional Long Short-Term Memory (BiLSTM) to learn the transition rules among sleep stages. A hierarchical neural network [28] implements comprehensive feature learning stage and sequence learning stage, respectively. Additionally, with the development of attention mechanisms, a deep Bi-directional RNN with attention mechanism is utilized for single-channel sleep staging [29].

Although CNN and RNN models achieve high accuracy, their limitation is that the model's input must be grid data ignoring the connection among brain regions. As different brain regions are not in the Euclidean space, grid data may not be the optimal data representation. Hence, the graph is the most appropriate data structure. GraphSleepNet [15] is proposed to utilize graph neural network to model functional connectivity brain network to achieve the SOTA performance. However, it only considers the spatial functional connectivity, and to a certain extent ignores the spatial proximity of brain regions.

Some previous researchers attempt to solve the subject difference problem found in physiological signals. Transfer learning methods are applied to improve the robustness of deep learning models for individual differences [17], [18]. For example, MetaSleepLearner [18] based on model-agnostic meta-learning is proposed to overcome the subject difference problem by training in the source domain and fine-tuning in the target domain. Although the existing transfer learning methods for sleep stage classification can achieve improved results, almost all existing work needs to fine-tune the pre-trained model for sleep stage classification. That is, these models require additional fine-tuning operations using part of the labeled data in the target domain. Therefore, these transfer learning methods that need to be fine-tuned are only suitable for the specific subject's personalized model. In this case, whenever a new subject needs to be evaluated, data must be collected again and the existing model must be fine-tuned again. Therefore, for clinical systems that need to be adapted to unknown subjects, fine-tuning operations may become inefficient.

III.   PRELIMINARIES

A.   Sleep Stage

Polysomnography (PSG) is usually employed for recording physiological signals during sleep in clinical medicine. The PSG is segmented into 30-second epochs for sleep stage classification. Sleep experts usually classify sleep epochs into different stages based on the sleep staging standard. Specifically, according to the AASM sleep staging standard, the human sleep process can be divided into three main parts: Wakefulness (Wake), rapid eye movement (REM), and non-rapid eye movement (NREM). Furthermore, the NREM can be

subdivided into three parts: N1 stage, N2 stage, and N3 stage. In general, sleep experts directly divide the sleep state into 5 different classes (Wake, N1, N2, N3, and REM).

B. Sleep Brain Network

A sleep brain network is defined as a graph $G = (V, E, A)$, where V represents the set of vertices and each vertex in the network represents an electrode on brain; $|V| = N$ is the number of vertices in sleep brain network; E denotes the set of edges and indicates the connection between vertices; A denotes the adjacency matrix of sleep brain network G. As presented in Figure 2, G_t^{FC} represents sleep brain network constructed from the functional connectivity and G_t^{DC} represents sleep brain network constructed from spatial distance. And a 30s EEG signal sequence S_t (called a sleep epoch) is transformed into G_t^{FC} and G_t^{DC}.

C. Sleep Feature Matrix

The sleep feature matrix is the input of the graph neural network. We define the raw signals sequence as $S = (S_1, S_2, ..., S_L) \in \mathbb{R}^{N \times T_s \times L}$, where L is the number of sleep epochs, T_s represents the time series length of each sleep epoch $S_i \in S (i \in \{1, 2, \cdots, L\})$. For each sleep epoch S_i, we extract the node feature by using a feature extraction network in Supplementary Material S.1 and define each epoch S_i's feature matrix $X_i = \left(x_1^i, x_2^i, ..., x_N^i \right)^\mathrm{T} \in \mathbb{R}^{N \times F_d}$, where $x_n^i \in \mathbb{R}^{F_d} (n \in \left\{ 1, 2, \cdots, N \right\})$ represents F_d features of node n at epoch i.

D. Sleep Stage Classification Problem

The research goal is to learn the mapping relationship between the encoded signals and sleep stage classes. The problem of sleep stage classification is defined as: given $\mathcal{S} = (S_{i-d}, ..., S_i, ..., S_{i+d}) \in \mathbb{R}^{N \times T_s \times T_n}$ identify the current sleep stage y, where $\mathcal{S}$ represents the temporal context of S_i, y denotes the S_i's sleep stage class label, and $T_n = 2d + 1$ is the number of sleep brain networks, where $d \in \mathbb{N}^+$ is temporal context. Specifically, in order to identify the sleep stage of the current sleep epoch S_i, we utilize its previous d epochs and following d epochs as the context. For each epoch, we construct G^{DC} and G^{FC} respectively, and they are employed as the input of our model to identify the sleep stage y of the current sleep epoch.

IV. MULTI-VIEW SPATIAL-TEMPORAL GCN

The overall architecture of the proposed model is exhibited in Figure 1. We summarize four key ideas of the proposed MSTGCN model: 1) Construct multiple views of the brain connection to fully indicate the spatial information of the brain. 2) Combine spatial graph convolution and temporal convolution to extract both spatial and temporal features. 3) Employ a spatial-temporal attention mechanism to automatically pay more attention to valuable spatial-temporal information. 4) Integrate domain generalization and spatial-temporal GCN in a unified framework to extract subject-invariant sleep features. The overall architecture is designed to accurately identify sleep stages.

A. Multi-View on Brain Graph

In this section, we introduce two different views from the brain graph: the functional connectivity-based brain graph and spatial distance-based brain graph. Different views reflect different spatial relationships of the brain. Specifically, the functional connectivity-based brain graph can present the collaboration of different brain regions in space. The actual physical locations of these brain regions may not be adjacent. However, existing neuroscience studies have presented physically adjacent brain regions also interact. Therefore, these two views on brain have a certain degree of complementarity and can fully demonstrate the spatial relationship of the brain.

1). Functional Connectivity-Based Brain Graph: Functional connectivity is usually constructed based on correlations or dependencies among physiological signals [30]. Pearson Correlation Coefficient (PCC) [31] and Mutual Information (MI) [32] are two common methods to determine the functional connectivity of the brain. Due to the limited understanding of the brain, it is still challenging to determine a suitable graph structure in advance for sleep stage classification. Hence, we propose a data-driven graph generation for functional connectivity. This data-driven approach constructs the functional connectivity graphs adaptively for different sleep stages based on the feature correlation between nodes as displayed in Figure 3. We define a non-negative function $A_{mn}^{FC} = g(x_m, x_n)(m, n \in \{1, 2, \cdots, N\})$ to represent the functional connectivity between nodes x_m and x_n based on the input feature matrix $X_i = \left(x_1^i, x_2^i, ..., x_N^i\right)^{\mathrm{T}} \in \mathbb{R}^{N \times F_d}$. $g(x_m, x_n)$ is implemented through a layer neural network, which has the learnable weight vector $w = \left(w_1, w_2, ..., w_{F_d}\right)^{\mathrm{T}} \in \mathbb{R}^{F_d \times 1}$. The learned graph structure (adjacency matrix) A^{FC} is defined as:

$$A_{mn}^{FC} = g(x_m, x_n) = \frac{\exp\left(\mathrm{ReLU}\left(w^{\mathrm{T}}|x_m - x_n|\right)\right)}{\sum_{n=1}^{N} \exp\left(\mathrm{ReLU}\left(w^{\mathrm{T}}|x_m - x_n|\right)\right)} \tag{1}$$

where rectified linear unit (ReLU) is an activation function to guarantee that A_{mn}^{FC} is non-negative. The softmax operation normalizes each row of A^{FC}. The weight vector w is updated by minimizing the following loss function,

$$\mathscr{L}_{\mathrm{graph\_learning}} = \sum_{m, n=1}^{N} \|x_m - x_n\|_2^2 A_{mn}^{FC} + \lambda \|A^{FC}\|_F^2 \tag{2}$$

That is, the larger distance $\|x_m - x_n\|_2$ between x_m and x_n, the smaller A_{mn}^{FC} is. Due to the brain connection structure is not a fully connected graph, we utilize the second term in the loss function to control the sparsity of graph A^{FC}, where $\lambda = 0.001$ is a regularization parameter.

The proposed graph generation mechanism automatically constructs the neighborhood connection of the nodes. To avoid the trivial solution (i.e., $w = (0, 0, \cdots, 0)$), which is

due to minimizing the above loss function $\mathscr{L}_{\text{graph\_learning}}$ independently, we utilize it as a regularized term to form the loss function.

2) Spatial Distance-Based Brain Graph: Previous studies have presented that adjacent brain regions affect each other and the strength of the impact is inversely proportional to the actual physical distance [16]. That is, the closer the distance between brain regions, the greater the impact. Therefore, we construct a spatial distance-based brain graph for sleep stage classification, as illustrated in Figure 4.

B. Spatial-Temporal Attention

The attention mechanism is often utilized to automatically extract the most relevant information. In this study, we employ a spatial-temporal attention mechanism [15] to capture valuable spatial-temporal information on the sleep brain network. The spatial-temporal attention mechanism contains spatial attention and temporal attention.

1) Spatial Attention: In the spatial dimension, different regions have different effects on the sleep stage which are dynamically changing during sleep. To automatically extract the attentive spatial dynamics, we utilize a spatial attention mechanism, which is defined as follows (take the spatial attention based on the functional connectivity view as an example):

$$P = V_p \cdot \sigma\left(\left(\mathscr{X}^{(l-1)}Z_1\right)Z_2\left(Z_3\mathscr{X}^{(l-1)}\right)^{\mathrm{T}} + b_p\right) \tag{3}$$

$$P'_{m,n} = \text{softmax}\left(P_{m,n}\right) \tag{4}$$

where $\mathscr{X}^{(l-1)} = \left(X_1, X_2, ..., X_{T_{l-1}}\right) \in \mathbb{R}^{N \times C_{l-1} \times T_{l-1}}$ is the l-th layer's input. C_{l-1} represents neural network channel's number of each node, i.e., $l=1$, $C_0 = F_d \cdot T_{l-1}$ denotes the l-th layer's temporal dimension. $V_p, b_p \in \mathbb{R}^{N \times N}$, $Z_1 \in \mathbb{R}^{T_{l-1}}$, $Z_2 \in \mathbb{R}^{C_{l-1} \times T_{l-1}}$, $Z_3 \in \mathbb{R}^{C_{l-1}}$ are learnable parameters, σ denotes the sigmoid activation function. P represents spatial attention matrix, which is dynamically computed by current layer's input. $P_{m,n}$ represents the correlation between node m and n. The softmax operation is utilized to normalize the attention matrix P. In the proposed model, when the graph convolution is performed, the learned adjacency matrix A^{FC} and spatial attention matrix P can dynamically adjust the update of nodes.

2) Temporal Attention: In the temporal dimension, there are correlations among neighboring sleep stages, and the correlations vary in different situations. Therefore, a temporal attention mechanism is utilized to capture dynamic temporal information among sleep brain networks.

The temporal attention mechanism is defined as follows:

$$Q = V_q \cdot \sigma\left(\left(\left(\mathscr{X}^{(l-1)}\right)^{\mathrm{T}}M_1\right)M_2\left(M_3\mathscr{X}^{(l-1)}\right) + b_q\right) \tag{5}$$

$$Q'_{u,v} = \text{softmax}(Q_{u,v}) \tag{6}$$

where $V_g, b_q \in \mathbb{R}^{T_{l-1} \times T_{l-1}}$, $M_1 \in \mathbb{R}^N$, $M_2 \in \mathbb{R}^{C_{l-1} \times N}$, $M_3 \in \mathbb{R}^{C_{l-1}}$ denotes learnable parameters. $Q_{m,n}$ denotes the strength of correlation between sleep brain network G_u and G_v. Finally, the softmax operation is utilized to normalize the attention matrix Q. The input of the MST-GCN is tuned by the temporal attention: $\widehat{\mathcal{X}}^{(l-1)} = (\widehat{X}_1, \widehat{X}_2, ..., \widehat{X}_{T_{l-1}}) = (X_1, X_2, ..., X_{T_{l-1}})Q' \in \mathbb{R}^{N \times C_{l-1} \times T_{l-1}}$ to pay more attention to informative temporal information.

C. Spatial-Temporal Graph Convolution

Spatial-temporal graph convolution is a combination of spatial graph convolution and standard temporal convolution, which is utilized to extract both spatial and temporal features. The spatial features are extracted by aggregating information from neighbor nodes for each sleep brain network and the temporal features are captured by exploiting temporal dependencies from neighbor sleep stages.

1) Spatial Graph Convolution: We employ graph convolution based on spectral graph theory to extract spatial features in the spatial dimension. For each sleep stage to be identified, the adjacency matrices A^{FC} and A^{DC} are provided for graph convolution. In addition, we employ the Chebyshev expansion of graph Laplacian to reduce computational complexity. Chebyshev graph convolution [33] using the $K-1$ order polynomials is defined as:

$$g_\theta *_G x = g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k T_k(\widetilde{L})x \tag{7}$$

where g_θ denotes the convolution kernel, $*G$ denotes the graph convolution operation, $\theta \in \mathbb{R}^K$ is a vector of Chebyshev coefficients and x is the input data. $L = D - A$ is Laplacian matrix, where $D \in \mathbb{R}^{N \times N}$ is degree matrix. $\widetilde{L} = \frac{2}{\lambda_{\max}}L - I_N$, where $\lambda_{\max}$ is Laplacian matrix's maximum eigenvalue and I_N is an identity matrix. $T_k(x)$ is the Chebyshev polynomials recursively.

The information of neighboring 0 to $K-1$ order neighbors centered at each node is extracted via the approximate expansion of Chebyshev polynomial.

We generalize the above definition to the nodes with multiple neural network channels. The l-th layer's input is $\widehat{\mathcal{X}}^{(l-1)} = (\widehat{X}_1, \widehat{X}_2, ..., \widehat{X}_{T_{l-1}}) \in \mathbb{R}^{N \times C_{l-1} \times T_{l-1}}$, where C_{l-1} represents neural network channel's number of each node, T_{l-1} denotes the l-th layer's temporal dimension. For each $\widehat{X}_i$, we obtain $g_\theta *_G \widehat{X}_i$ by using C_l filters on $\widehat{X}_i$, where $\Theta = (\Theta_1, \Theta_2, ..., \Theta_{C_l}) \in \mathbb{R}^{K \times C_{l-1} \times C_l}$ is the convolution kernel parameter [33]. Hence, the information of the $0 \sim K-1$ order neighbors is aggregated to each node.

2) Temporal Convolution: To capture the sleep transition rules, which are utilized by sleep experts to classify the current sleep stage in combination with neighboring sleep stages, we employ CNN to perform convolution operation in the temporal dimension. Specifically, after graph convolution operation has sufficiently extracted the spatial features from each sleep brain network, we implement a standard 2D convolution layer to extract the temporal context information of the current sleep stage. The temporal convolution operation on the *l*-th layer is defined as:

$$\mathcal{X}^{(l)} = \text{ReLU}\left(\Phi * \left(\text{ReLU}\left(g_\theta *_G \widehat{\mathcal{X}}^{(l-1)}\right)\right)\right) \in \mathbb{R}^{N \times C_l \times T_l} \tag{8}$$

where ReLU is the activation function, Φ denotes the convolution kernel's parameters, * denotes the standard convolution operation.

After the multi-view ST-GCN extracts a large number of features, we employ the concatenate operation to perform feature fusion on $\mathcal{X}^{FC}$ and $\mathcal{X}^{DC}$:

$$\mathcal{X} = \mathcal{X}^{FC} \parallel \mathcal{X}^{DC} \tag{9}$$

where $\mathcal{X}^{FC}$, $\mathcal{X}^{DC}$ represent the features respectively extracted from functional connectivity and spatial distance based view, $\parallel$ is the concatenate operation.

D. Domain Generalization

In order to reduce the influence of individual differences, we exploit an adversarial domain generalization method to enhance the robustness of our model. Figure 5 presents the intuitive idea of the adversarial domain generalization. Specifically, this method aims to make it impossible to distinguish which source domain the sample data originated from during model training. At the same time, it aims to improve the sleep stage classification performance as much as possible. This means that all subjects' common features (subject-invariant features) related to sleep stage classification are extracted. For example, the model cannot distinguish that the samples of Domain 1 are the data belonging to its own domain, but it can still accurately identify the sleep stages. This presents that the model did not learn personalized features (F-1) belonging to Domain 1, but some common features related to sleep stage classification. In fact, previous studies have presented the advantages of adversarial domain generalization [34], and theoretically this method aligns the marginal distribution of different domains. Specifically, domain generalization includes three parts: feature extractor $\mathcal{G}_f$, domain classifier $\mathcal{G}_d$ and label predictor $\mathcal{G}_y$. The feature extractor $\mathcal{G}_f$ maps the input data to a domain-invariant feature space,

$$\mathcal{G}_f(\mathbf{X}; \theta_f) = \mathcal{X} \tag{10}$$

where $\mathbf{X}$ is the input feature matrix, θ_f is the trainable parameter and $\mathcal{X}$ is the transferred feature matrix.

The transferred features are put into label predictor $\mathcal{G}_y$ and domain classifier $\mathcal{G}_d$ with softmax function:

$$\widehat{y}_i = \frac{\exp(w_y \mathscr{X}_i + b_y)}{\sum_{i=1}^{N} \exp(w_y \mathscr{X}_i + b_y)} \tag{11}$$

$$\widehat{d}_i = \frac{\exp(w_d \mathscr{X}_i + b_d)}{\sum_{i=1}^{N} \exp(w_d \mathscr{X}_i + b_d)} \tag{12}$$

where $\mathscr{X}_i$ denotes the transferred features of sample i. $\widehat{y}_i$ and $\widehat{d}_i$ are the predicted results of $\mathscr{G}_y$ and $\mathscr{G}_d$, respectively. Both of the $\mathscr{G}_y$ and $\mathscr{G}_d$ are multi-class classifier, we employ the cross entropy as the loss function:

$$\mathscr{L}_y = -\frac{1}{L} \sum_{i=1}^{L} \sum_{r=1}^{R_y} y_{i,r} \log \widehat{y}_{i,r} \tag{13}$$

$$\mathscr{L}_d = -\frac{1}{L} \sum_{i=1}^{L} \sum_{r=1}^{R_d} d_{i,r} \log \widehat{d}_{i,r} \tag{14}$$

where $\mathscr{L}_y$ is the cross entropy loss function of the multi-classification task, L denotes the number of samples, R_y and R_d denote the number of classes and the number of domains, respectively. y is the true label and $\widehat{y}$ is the value predicted by the model. d is the true domain and $\widehat{d}$ is the value predicted by the model.

Besides, a special layer called Gradient Reversal Layer (GRL) is implemented between feature extractor $\mathscr{G}_f$ and domain classifier $\mathscr{G}_d$ to form an adversarial relationship [35]. Compared with other methods that usually require training classifier and discriminator in separate steps, GRL can integrate feature learning and domain generalization in a unified framework and execute back-propagation algorithms. The optimization process is defined as:

$$\begin{aligned} \left(\widehat{\theta}_f, \widehat{\theta}_y\right) &= \arg\min_{\theta_f, \theta_y} \mathscr{L}\left(\theta_f, \theta_y, \widehat{\theta}_d\right) \\ \left(\widehat{\theta}_d\right) &= \arg\max_{\theta_d} \mathscr{L}\left(\widehat{\theta}_f, \widehat{\theta}_y, \theta_d\right) \end{aligned} \tag{15}$$

where θ_d, θ_y are the parameters to minimize the loss of $\mathscr{G}_d$ and $\mathscr{G}_y$, respectively. θ_f is the parameters of $\mathscr{G}_f$ to minimize the loss of $\mathscr{G}_y$ and maximize the loss of $\mathscr{G}_d$ at the same time. The aims of feature extractor $\mathscr{G}_f$ and domain classifier $\mathscr{G}_d$ are exact opposite. The feature extractor $\mathscr{G}_f$ aims to make the domain classifier $\mathscr{G}_d$ can't classify the right domain and the domain classifier $\mathscr{G}_d$ aims to correctly classify the domain that the data comes from.

The whole loss function of the domain generalization is defined as:

$$\mathscr{L}_{\text{DG}} = -\frac{1}{L}\sum_{i=1}^{L}\sum_{r=1}^{R_y} y_{i,r}\log\hat{y}_{i,r} + \beta\frac{1}{L}\sum_{i=1}^{L}\sum_{r=1}^{R_d} d_{i,r}\log\hat{d}_{i,r} \qquad (16)$$

By optimizing the loss function, the feature extractor $\mathscr{G}_f$ can achieve the goal of finding the domain-invariant feature space.

V. EXPERIMENTS AND DISCUSSIONS

A. Dataset and Experiment Settings

Two publicly available datasets are employed in our experiments: 1) **ISRUC-S3 dataset** [36] contains 10 healthy subjects (9 male and 1 female). Each recording contains 6 EEG channels, 2 EOG channels, 3 EMG channels, and 1 ECG channel. In addition, the experts classify these PSG recordings into five sleep stages according to AASM standard [1]. 2) **MASS-SS3 dataset** [37] contains 62 healthy subjects (28 male and 34 female). Each recording contains 20 EEG channels, 2 EOG channels, 3 EMG channels, and 1 ECG channel.

We compare our MSTGCN with 7 baselines, which are described in detail in Supplementary Material S.3. For a fair comparison, we employ the same experimental settings for all models. Specifically, we employ 10-fold cross-validation and 31-fold cross-validation to evaluate the performance of all models on ISRUC-S3 dataset and MASS-SS3 dataset, respectively. In addition, we adopt the subject-independent strategy for cross-validation. We implement the proposed model using TensorFlow. In addition, the code is released on Github.[1]

B. Comparison With the State-of-the-Art Methods

We compare the proposed model with the other baseline models for sleep stage classification on the ISRUC-S3 and MASS-SS3 as presented in Table I and Table II. The results present that our proposed model outperforms the baseline methods on multiple overall metrics (overall Accuracy, F1-score, and Kappa) for ISRUC-S3 and MASS-SS3. Specifically, the traditional machine learning methods (SVM and RF) cannot learn the complex spatial or temporal features well. However, existing deep learning models such as CNN and RNN [7], [11]–[13] can directly extract the spatial or temporal features. Therefore, their performance is better than the traditional machine learning methods.

Although CNN and RNN achieve high accuracy, their limitation is that the model's input must be grid data ignoring the connection among brain regions. Due to brain regions are in non-Euclidean space, graph is the most appropriate data structure to indicate the connections. Therefore, the proposed model and ST-GCN can often achieve optimal or suboptimal overall results, especially on the MASS-SS3 dataset. In addition, the proposed model extracts both spatial and temporal features based on multi-view brain graphs and

---

[1] https://github.com/ziyujia/MSTGCN

integrates domain generalization to learn subject-invariant features. Hence, the proposed model achieves the state-of-the-art performance.

For different sleep stages, MSTGCN can accurately identify most of the corresponding stages. Specifically, in the ISRUC-S3 dataset, the classification accuracy of Wake and N3 stages is the highest. In the MASS-SS3 dataset, the classification accuracy of the REM and N2 stages is the highest. However, the classification performance of the N1 stage does not meet expectations on the two datasets, like other baseline models. It may be because the N1 stage is a transitional period between the Wake stage and the N2 stage, and the sample number of N1 stage is relatively small. Therefore, as Figure S.2 in Supplementary Material shows, N1 stage is mistakenly divided into other sleep stages, such as Wake stage and N2 stage. Nevertheless, the classification performance of MSTGCN for the N1 stage is still higher than most baseline models. Table II presents that MSTGCN has the highest F1-score for N1 stage on the MASS-SS3 dataset, which is 4% higher than the sub-optimal result.

C. Experimental Analysis and Discussion

1) Ablation Experiment: To validate the effect of each module in our model, we design some variant models. First, we use the spatial graph convolution with spatial distance brain graph as the basic model to gradually stack the remaining modules to form a whole branch. Then, we add another whole ST-GCN branch with functional connectivity brain graph to form a multi-view ST-GCN. Finally, we integrate the domain generalization method to form the proposed model. The specific process is described as follows:

- *variant a (Spatial Graph Convolution (Base Model)):* We utilize a spatial graph convolution network with spatial distance brain graph as the base model.

- *variant b (+ Temporal Convolution):* We add temporal convolution to form a spatial-temporal graph convolution network.

- *variant c (+ Attention Mechanism):* We add attention mechanism both on spatial and temporal dimension.

- *variant d (+ Multi-view Fusion (Add Another View)):* We add another whole ST-GCN branch based on the functional connectivity brain graph to form a multi-view ST-GCN.

- *variant e (+ Domain Generalization):* A multi-view ST-GCN with domain generalization (our MSTGCN).

Figure 6 presents that the key modules in our model are effective for sleep stage classification, especially variant *c*, variant *d*, and variant *e*. Specifically, the attention mechanism helps to capture valuable spatial-temporal features to improve the classification performance of our model. The designed multi-view on brain provides complementary information for sleep stage classification. In addition, domain generalization is integrated into the multi-view ST-GCN to extract subject-invariant features, which helps to improve the model generalization. In summary, the ablation experiment presents the effectiveness of each module in our model.

2) Adaptive Functional Connectivity Graph: To further investigate the effectiveness of the adaptive functional connectivity graph learning, we design five fixed functional connectivity graphs to compare with it. These graphs are defined as different adjacency matrices. The last three graphs are constructed by functional connectivity methods commonly found in neuroscience.

- Fully Connected Adjacency Matrix: A matrix whose elements are all 1. It represents that there are all connections among all nodes and each node also has self-connection in the graph.

- *K*-Nearest Neighbor (KNN) Adjacency Matrix [38]: A matrix, which represents a *k*-nearest neighbor graph. That is, each node has *k* neighbor nodes.

- Pearson Correlation Coefficient (PCC) Adjacency Matrix [31]: A matrix generated by the pearson correlation coefficient between each pair of nodes.

- Phase Locking Value (PLV) Adjacency Matrix [39]: A matrix generated by the PLV method between each pair of nodes.

- Mutual Information (MI) Adjacency Matrix [32]: A matrix generated by measuring the mutual dependence between each pair of nodes.

Figure 7 illustrates that the adaptive (learned) adjacency matrix achieves the highest accuracy for sleep stage classification. In addition, the adjacency matrix combined with prior neuroscience knowledge also achieves a suboptimal effect, such as the PCC, PLV, and MI adjacency matrix. The fully connected adjacency matrix does not work well because the brain network is not a fully connected graph. In general, the adjacency matrix can significantly affect the classification performance. The proposed adaptive functional connectivity graph for classification tasks is superior to the fixed functional connectivity graphs.

To present the interpretability of the adaptive functional connectivity graph, we visualize the brain adjacency matrices obtained by adaptive learning for different sleep stages. These matrices reflect the brain functional connectivity in different sleep stages as illustrated in Figure 8. Specifically, there are more functional connectivity in the Wake stage and N1 stage. On the contrary, the functional connectivity of the N3 stage is the least. These findings are consistent with existing neuroscience research [40], [41]. N3 stage is a typical deep sleep period, and the brain is usually in an inactive stage. In contrast, the N1 stage is a light sleep period, and the brain is relatively active. Therefore, the functional connectivity of the brain in the N1 stage is relatively complicated.

3) Attention Mechanism: To explore the interpretability of the attention mechanism, first we visualize the learned weight of temporal attention mechanism to indicate the importance of different sleep epochs for classification. The higher the weight, the higher the degree of attention. Figure 9 illustrates that the weight of the current sleep stage *T* is the largest. Previous and following sleep epochs received similar but lower attention. That is, this stage has received the most attention, which is consistent with the AASM sleep standard [1]. In fact, sleep experts mainly judge the current sleep stage type based on the characteristics of the current sleep state and appropriately refer to the adjacent sleep state.

Therefore, the temporal attention mechanism has learned expert knowledge to a certain extent.

In addition, we also visualize the learned weight of spatial attention mechanism for EEG channels. Figure 10 illustrates that our model pays different attention to EEG channels in different sleep stages, which may caused by the EEG patterns of different sleep stages are different. The attention weights of F3 and F4 are always the lowest. In contrast, the attention weights of C3 and C4 have always been the highest for different sleep stages. The results indicate that C3 and C4 may be the most informative EEG channels for sleep stage classification. Generally, the C3 and C4 channels are located in the middle of the scalp, which may have richer EEG information and be less affected by external factors.

VI. CONCLUSION

In this paper, we propose a novel deep graph neural network MSTGCN for sleep stage classification. In MSTGCN, we propose effective approaches in modeling the dynamics of sleep data along both the spatial and temporal dimensions, as well as considering the subject differences in sleep data. Specifically, we design different brain views based on the functional connectivity and physical distance proximity of the brain. The complementarity of different views provides rich spatial topology information. We develop a spatial-temporal graph convolution with attention mechanism to simultaneously capture the most relevant spatial-temporal features for sleep stage classification. Moreover, to extract subject-invariant sleep features, we integrate domain generalization and spatial-temporal graph convolutional networks into a unified framework. Experiments on two public sleep datasets demonstrate MSTGCN achieves the state-of-the-art performance. Finally, our proposed approach provides a general-framework for multivariate physiological time series.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

[1]. Berry RB et al. , "Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events," J. Clin. Sleep Med, vol. 8, no. 5, pp. 597–619, 2012. [PubMed: 23066376]

[2]. Alickovic E. and Subasi A, "Ensemble SVM method for automatic sleep stage classification," IEEE Instrum. Meas, vol. 67, no. 6, pp. 1258–1265, 6. 2018.

[3]. Memar P. and Faradji F, "A novel multi-class EEG-based sleep stage classification system," IEEE Trans. Neural Syst. Rehabil. Eng, vol. 26, no. 1, pp. 84–95, 1. 2018. [PubMed: 29324406]

[4]. LeCun Y, Bottou L, Bengio Y, and Haffner P, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 11. 1998.

[5]. Elman JL, "Finding structure in time," Cogn. Sci, vol. 14, no. 2, pp. 179–211, 1990.

[6]. Jia Z, Lin Y, Wang J, Wang X, Xie P, and Zhang Y, "SalientSleepNet: Multimodal salient wave detection network for sleep staging," 2021, arXiv:2105.13864. [Online]. Available: http://arxiv.org/abs/2105.13864

[7]. Chambon S, Galtier MN, Arnal PJ, Wainrib G, and Gramfort A, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," IEEE Trans. Neural Syst. Rehabil. Eng, vol. 26, no. 4, pp. 758–769, 4. 2018. [PubMed: 29641380]

[8]. Cai X, Jia Z, Tang M, and Zheng G, "BrainSleepNet: Learning multivariate EEG representation for automatic sleep staging," in Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM), Dec. 2020, pp. 976–979.

[9]. Jia Z, Cai X, Zheng G, Wang J, and Lin Y, "SleepPrintNet: A multivariate multimodal neural network based on physiological time-series for automatic sleep staging," IEEE Trans. Artif. Intell, vol. 1, no. 3, pp. 248–257, 12. 2020.

[10]. Jia Z, Lin Y, Zhang H, and Wang J, "Sleep stage classification model based ondeep convolutional neural network," J. ZheJiang Univ., Eng. Sci, vol. 54, no. 10, pp. 1899–1905, 2020.

[11]. Dong H, Supratak A, Pan W, Wu C, Matthews PM, and Guo Y, "Mixed neural network approach for temporal sleep stage classification," IEEE Trans. Neural Syst. Rehabil. Eng, vol. 26, no. 2, pp. 324–333, 2. 2018. [PubMed: 28767373]

[12]. Supratak A, Dong H, Wu C, and Guo Y, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," IEEE Trans. Neural Syst. Rehabil. Eng, vol. 25, no. 11, pp. 1998–2008, 11. 2017. [PubMed: 28678710]

[13]. Phan H, Andreotti F, Cooray N, Chén OY, and De Vos M, "SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," IEEE Trans. Neural Syst. Rehabil. Eng, vol. 27, no. 3, pp. 400–410, 3. 2019. [PubMed: 30716040]

[14]. Gopinath K, Desrosiers C, and Lombaert H, "Adaptive graph convolution pooling for brain surface analysis," in Proc. Int. Conf. Inf. Process. Med. Imag. Cham, Switzerland: Springer, 2019, pp. 86–98.

[15]. Jia Z. et al., "GraphSleepNet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification," in Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI), Jul. 2020, pp. 1324–1330.

[16]. Salvador R, Suckling J, Coleman MR, Pickard JD, Menon D, and Bullmore ED, "Neurophysiological architecture of functional magnetic resonance images of human brain," Cerebral Cortex, vol. 15, no. 9, pp. 1332–1342, 2005. [PubMed: 15635061]

[17]. Phan H. et al. , "Personalized automatic sleep staging with single-night data: A pilot study with KL-divergence regularization," 2020, arXiv:2004.11349. [Online]. Available: http://arxiv.org/abs/2004.11349

[18]. Banluesombatkul N. et al. , "MetaSleepLearner: A pilot study on fast adaptation of bio-signals-based sleep stage classifier to new individual subject using meta-learning," IEEE J. Biomed. Health Informat, vol. 25, no. 6, pp. 1949–1963, 6. 2021.

[19]. Sokolovsky M, Guerrero F, Paisarnsrisomsuk S, Ruiz C, and Alvarez SA, "Deep learning for automated feature discovery and classification of sleep stages," IEEE/ACM Trans. Comput. Biol. Bioinf, vol. 17, no. 6, pp. 1835–1845, 11. 2020.

[20]. Jia Z, Lin Y, Liu Y, Jiao Z, and Wang J, "Refined nonuniform embedding for coupling detection in multivariate time series," Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top, vol. 101, no. 6, 6. 2020, Art. no. 062113.

[21]. Jia Z, Lin Y, Jiao Z, Ma Y, and Wang J, "Detecting causality in multivariate time series via non-uniform embedding," Entropy, vol. 21, no. 12, p. 1233, 12. 2019.

[22]. Li Z, Wang J, Jia Z, and Lin Y, "Learning space-time-frequency representation with two-stream attention based 3D network for motor imagery classification," in Proc. IEEE Int. Conf. Data Mining (ICDM), Nov. 2020, pp. 1124–1129.

[23]. Ziyu J, Youfang L, Tianhang L, Kaixin Y, Xinwang Z, and Jing W, "Motor imagery classification based on multiscale feature extraction and squeeze-excitation model," J. Comput. Res. Develop, vol. 57, no. 12, p. 2481, 2020.

[24]. Jia Z, Lin Y, Wang J, Yang K, Liu T, and Zhang X, "MMCNN: A multi-branch multi-scale convolutional neural network for motor imagery classification," in Machine Learning and

Knowledge Discovery in Databases, Hutter F, Kersting K, Lijffijt J, and Valera I, Eds. Cham, Switzerland: Springer, 2021, pp. 736–751.

[25]. Jia Z, Lin Y, Cai X, Chen H, Gou H, and Wang J, "SST-EmotionNet: Spatial-spectral-temporal based attention 3D dense network for EEG emotion recognition," in Proc. 28th ACM Int. Conf. Multimedia, Oct. 2020, pp. 2909–2917.

[26]. Jia Z, Lin Y, Wang J, Feng Z, Xie X, and Chen C, "HetEmotionNet: Two-stream heterogeneous graph recurrent neural network for multimodal emotion recognition," 2021, arXiv:2108.03354. [Online]. Available: http://arxiv.org/abs/2108.03354

[27]. Zhang J. and Wu Y, "A new method for automatic sleep stage classification," IEEE Trans. Biomed. Circuits Syst, vol. 11, no. 5, pp. 1097–1110, 10. 2017. [PubMed: 28809709]

[28]. Sun C, Chen C, Li W, Fan J, and Chen W, "A hierarchical neural network for sleep stage classification based on comprehensive feature learning and multi-flow sequence learning," IEEE J. Biomed. Health Informat, vol. 24, no. 5, pp. 1351–1366, 5 2020.

[29]. Phan H, Andreotti F, Cooray N, Chén OY, and Vos MD, "Automatic sleep stage classification using single-channel EEG: Learning sequential features with attention-based recurrent neural networks," in Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2018, pp. 1452–1455.

[30]. Tagliazucchi E, von Wegner F, Morzelewski A, Borisov S, Jahnke K, and Laufs H, "Automatic sleep staging using fMRI functional connectivity data," NeuroImage, vol. 63, no. 1, pp. 63–72, 10. 2012. [PubMed: 22743197]

[31]. Pearson K. and Lee A, "On the laws of inheritance in man: I. Inheritance of physical characters," Biometrika, vol. 2, no. 4, pp. 357–462, 1903.

[32]. Danon L, Díaz-Guilera A, Duch J, and Arenas A, "Comparing community structure identification," J. Stat. Mech., Theory Exp, vol. 2005, no. 9, 9. 2005, Art. no. P09008.

[33]. Defferrard M, Bresson X, and Vandergheynst P, "Convolutional neural networks on graphs with fast localized spectral filtering," in Proc. Adv. Neural Inf. Process. Syst, 2016, pp. 3844–3852.

[34]. Li Y. et al., "Deep domain generalization via conditional invariant adversarial networks," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 624–639.

[35]. Ganin Y. and Lempitsky V, "Unsupervised domain adaptation by back-propagation," in Proc. Int. Conf. Mach. Learn., 2015, pp. 1180–1189.

[36]. Khalighi S, Sousa T, Santos JM, and Nunes U, "ISRUC-Sleep: A comprehensive public dataset for sleep researchers," Comput. Methods Programs Biomed, vol. 124, pp. 180–192, 2. 2016. [PubMed: 26589468]

[37]. O'Reilly C, Gosselin N, Carrier J, and Nielsen T, "Montreal archive of sleep studies: An open-access resource for instrument benchmarking and exploratory research," J. Sleep Res, vol. 23, no. 6, pp. 628–635, 2014. [PubMed: 24909981]

[38]. Jiang B, Ding C, Luo B, and Tang J, "Graph-Laplacian PCA: Closed-form solution and robustness," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 6. 2013, pp. 3492–3498.

[39]. Aydore S, Pantazis D, and Leahy RM, "A note on the phase locking value and its properties," NeuroImage, vol. 74, pp. 231–244, 7. 2013. [PubMed: 23435210]

[40]. Spoormaker VI et al. , "Development of a large-scale functional brain network during human non-rapid eye movement sleep," J. Neurosci, vol. 30, no. 34, pp. 11379–11387, 8. 2010. [PubMed: 20739559]

[41]. Larson-Prior LJ et al., "Modulation of the brain's functional network architecture in the transition from wake to sleep," in Progress in Brain Research, vol. 193. Amsterdam, The Netherlands: Elsevier, 2011, pp. 277–294. [PubMed: 21854969]
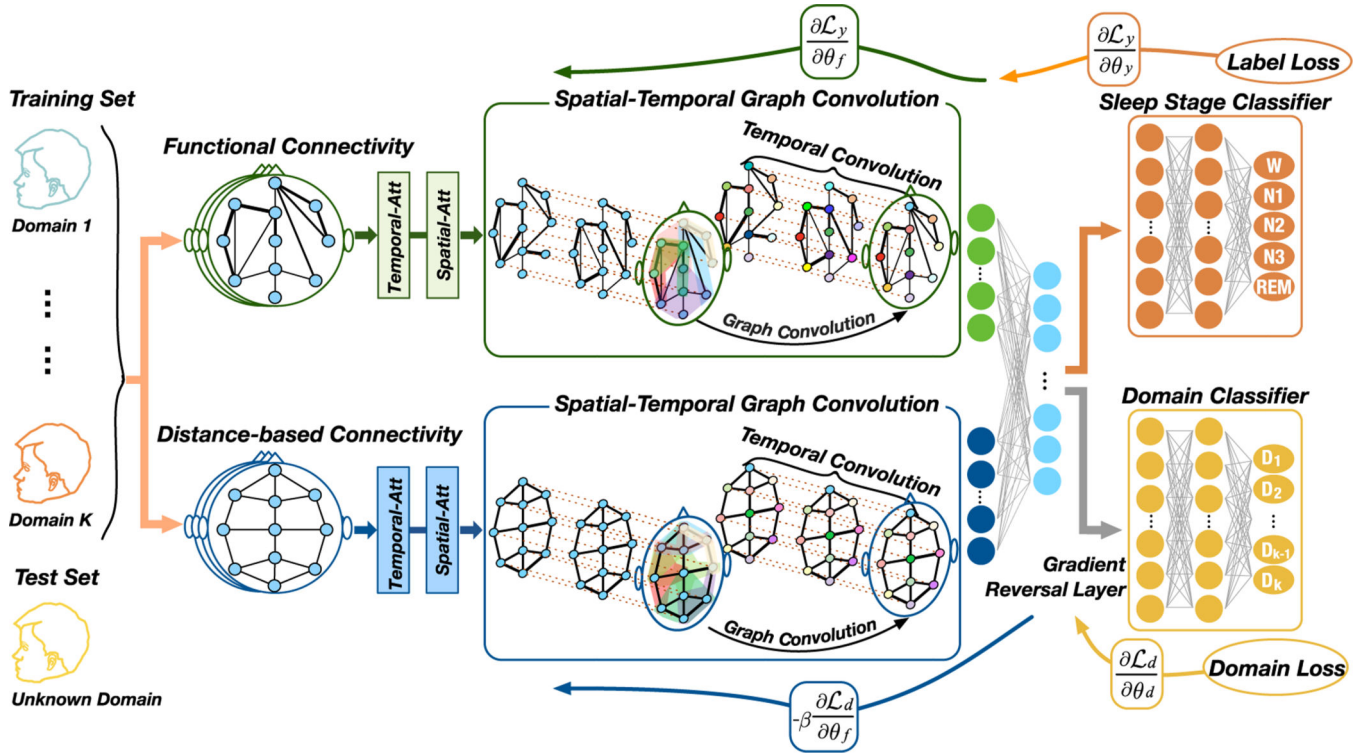
Fig. 1.
The overall architecture of the MSTGCN for sleep stage classification. First of all, two different views on the brain are constructed: the functional connectivity-based brain graph and spatial distance-based brain graph. Different views reflect different spatial relationships of the brain. Then, an attention based spatial-temporal graph convolution is designed for the most relevant spatial-temporal features for sleep stage classification. Finally, a domain generalization with the gradient reversal layer is implemented to improve the generalization of the model. In domain generalization, each subject in training set is treated as a specific source domain. The advantage of domain generalization over other transfer learning methods is that this method does not require any information (a small number of labeled samples or unlabeled sample data distribution) from the test set (called unknown domain or target domain). Therefore, domain generalization improves the generalization of the model and is more suitable for clinical systems applied by unknown users.
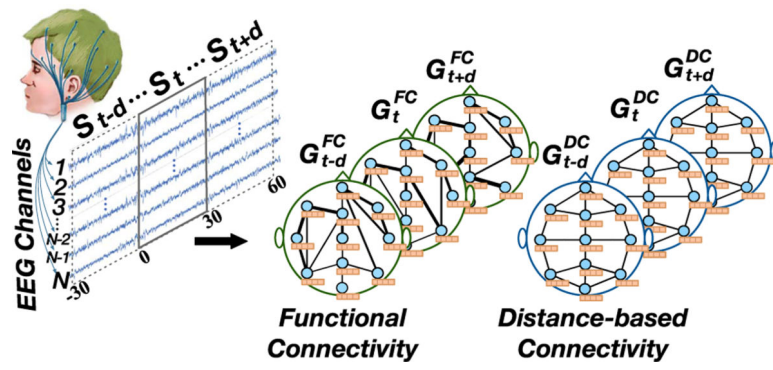
Fig. 2.
Multi-view sleep brain network. Left network is the functional connectivity-based network and right network is the spatial distance-based network.
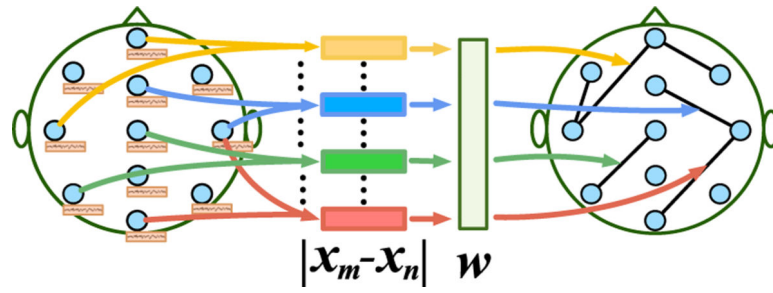
$$|x_m - x_n| \quad w$$

Fig. 3.

The adaptive sleep graph learning to generate functional connectivity for sleep stage classification. x_m and x_n represent the features of two nodes respectively, w is learnable weight. The more similar the node features, the greater the probability of establishing a connection.
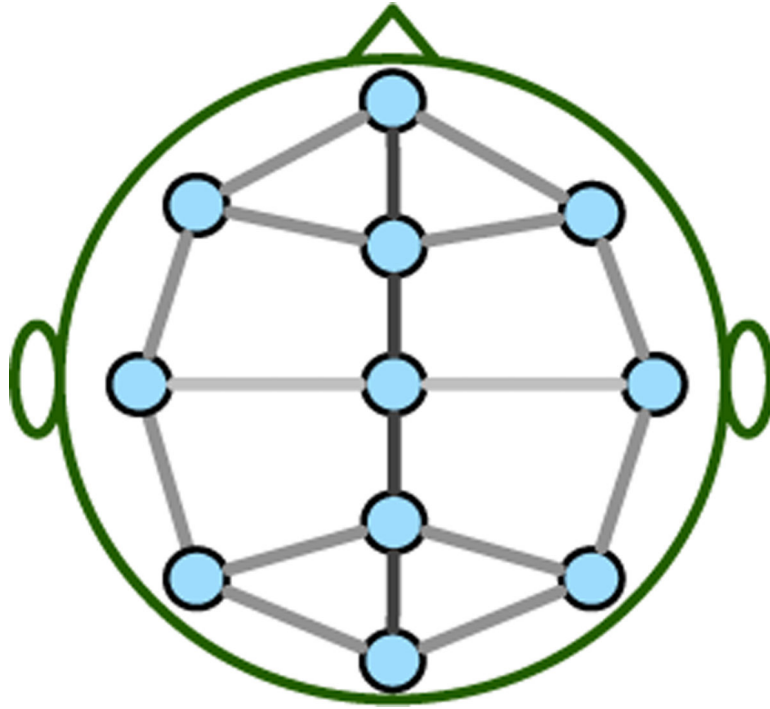
Fig. 4.
The spatial distance-based brain graph for sleep stage classification.
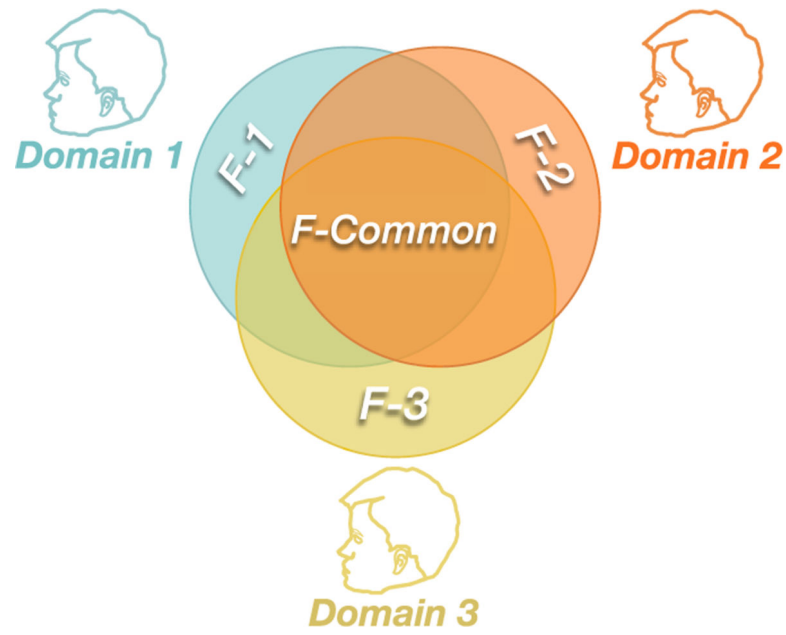
Fig. 5.

The intuitive idea of the adversarial domain generalization to extract subject-invariant features. Each subject is treated as a specific domain. F-Common means that all subjects have common features for sleep stage classification. F-1, F-2, and F-3 represent some of the subjects' unique features related to sleep stage classification. Domain generalization makes the model unable to distinguish which subject the sample comes from. At the same time, as much as possible to improve the model performance for sleep stage classification. This means that some unique features of subjects are not learned by the model, but some common subject-invariant features (F-Common) related to sleep stage classification are extracted. Therefore, the generalization of the model is improved through domain generalization.
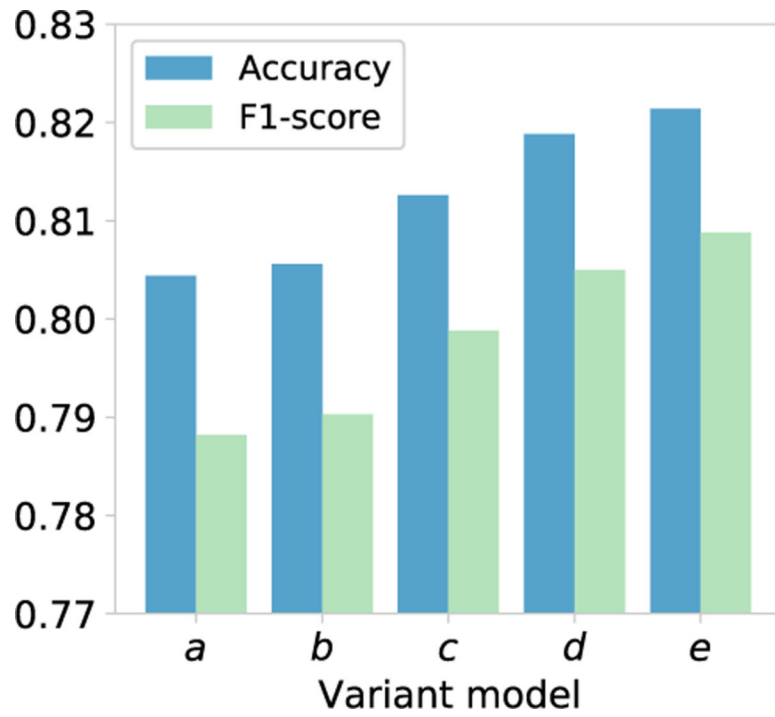
Fig. 6.
Comparison of the designed variant models to verify the effectiveness of different modules in MSTGCN.
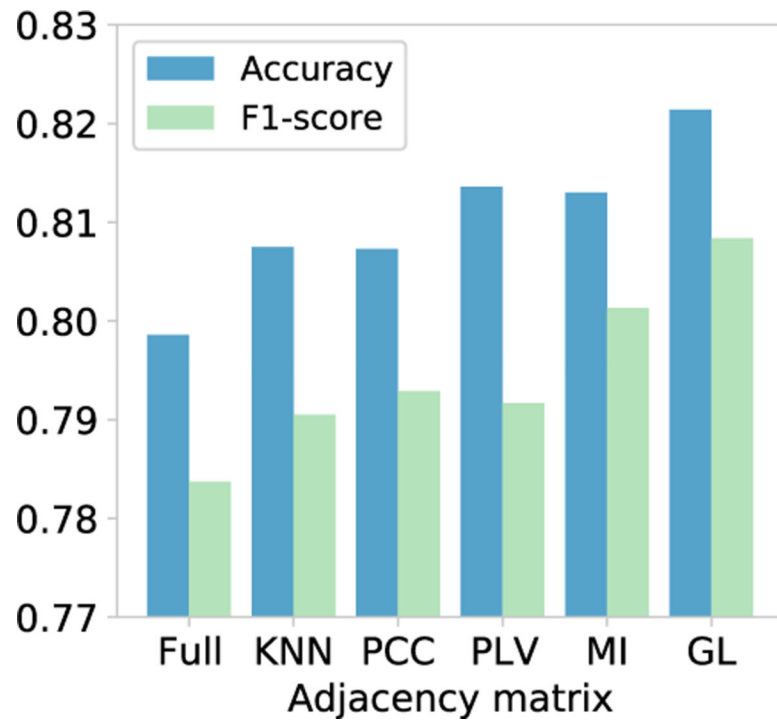
Fig. 7.
Comparison of different adjacency matrices. GL: the proposed Graph Learning approach for brain functional connectivity. Full: Fully Connected Adjacency Matrix; KNN: K-Nearest Neighbor Adjacency Matrix; PCC: Pearson Correlation Coefficient Adjacency Matrix; PLV: Phase Locking Value Adjacency Matrix; MI: Mutual Information Adjacency Matrix.
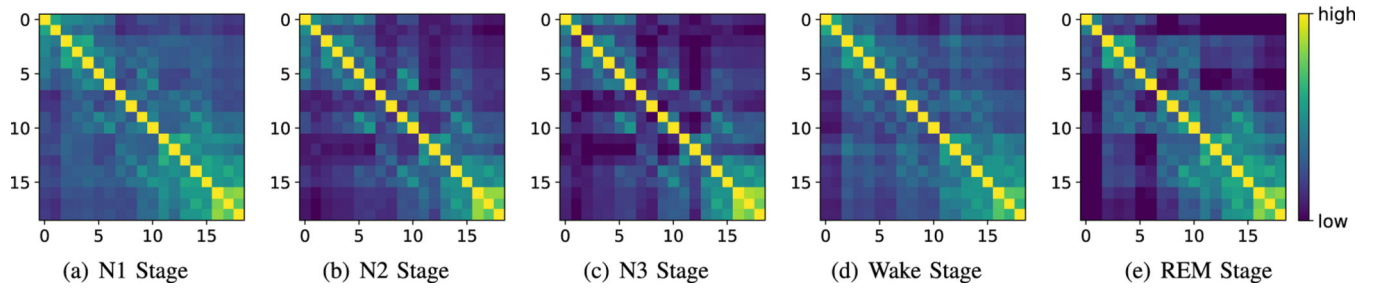
(a) N1 Stage (b) N2 Stage (c) N3 Stage (d) Wake Stage (e) REM Stage

Fig. 8.
The learned adjacency matrix visualization of five sleep stages (N1 Stage, N2 Stage, N3 Stage, Wake Stage, and REM Stage).
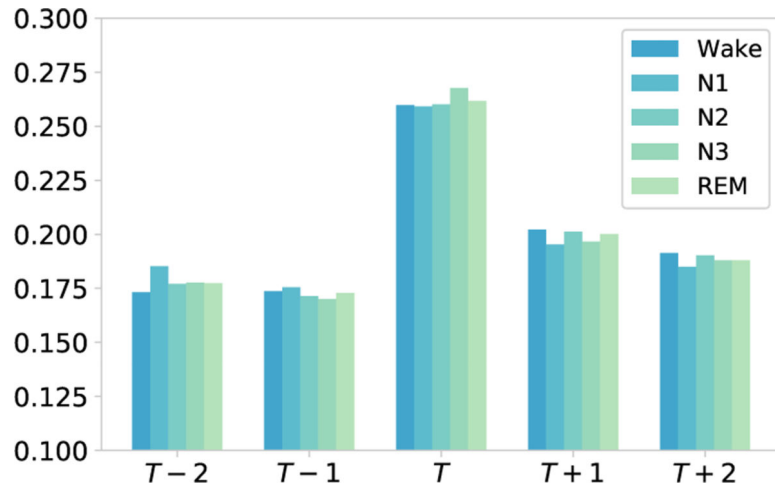
Fig. 9.
Temporal attention visualization. The current sleep stage T always keeps the most attention weights. The adjacent sleep stages keep some attention weights for this classification task.
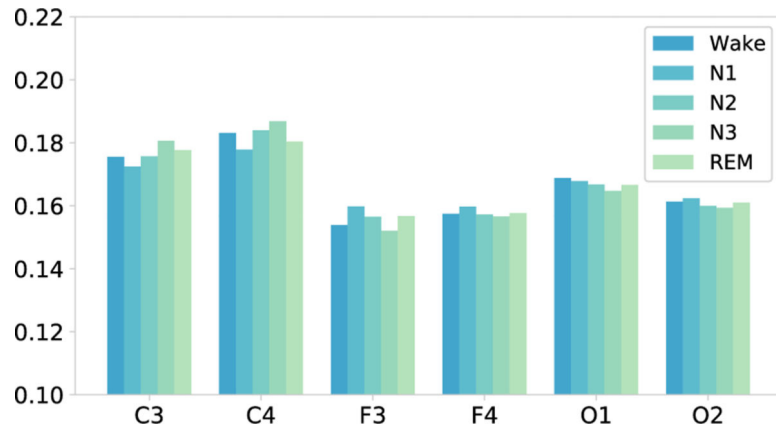
Fig. 10.
Spatial attention visualization to present the contribution of various EEG channels for sleep stage classification. The attention weights of C3 channel and C4 channel is always the highest for different sleep stages.

TABLE I

THE PERFORMANCE COMPARISON OF THE STATE-OF-THE-ART APPROACHES ON THE ISRUC-S3 DATASET

	Method	Overall results				F1-score for each class					
		Accuracy	F1-score	Kappa	Wake	N1	N2	N3	REM		
Alickovic et al. [2]	SVM	0.733	0.721	0.657	0.868	0.523	0.699	0.786	0.731		
Memar et al. [3]	RF	0.729	0.708	0.648	0.858	0.473	0.704	0.809	0.699		
Dong et al. [11]	MLP+LSTM	0.779	0.758	0.713	0.860	0.469	0.760	0.875	0.828		
Supratak et al. [12]	CNN+BiLSTM	0.788	0.779	0.730	0.887	**0.602**	0.746	0.858	0.802		
Chambon et al. [7]	CNN	0.781	0.768	0.720	0.870	0.550	0.760	0.851	0.809		
Phan et al. [13]	ARNN+RNN	0.789	0.763	0.725	0.836	0.439	0.793	0.879	**0.867**		
Jia et al. [15]	STGCN	0.799	0.787	0.741	0.878	0.574	0.776	0.864	0.841		
proposed model	**MSTGCN**	**0.821**	**0.808**	**0.769**	**0.894**	0.596	**0.806**	**0.890**	0.856		

*The bold result is the best result and the underlined result is the second best result.

TABLE II

THE PERFORMANCE COMPARISON OF THE STATE-OF-THE-ART APPROACHES ON THE MASS-SS3 DATASET

	Method	Overall results				F1-score for each class				
		Accuracy	F1-score	Kappa	Wake	N1	N2	N3	REM
Alickovic et al. [2]	SVM	0.779	0.688	0.659	0.801	0.339	0.843	0.645	0.813
Memar et al. [3]	RF	0.800	0.726	0.697	0.863	0.379	0.858	0.784	0.749
Dong et al. [11]	MLP+LSTM	0.859	0.805	-	0.846	0.563	0.907	0.848	0.861
Supratak et al. [12]	CNN+BiLSTM	0.862	0.817	0.800	0.873	0.598	0.903	0.815	0.893
Chambon et al. [7]	CNN	0.739	0.673	0.640	0.730	0.294	0.812	0.765	0.764
Phan et al. [13]	ARNN+RNN	0.871	0.833	0.815	-	-	-	-	-
Jia et al. [15]	STGCN	0.889	0.841	0.834	0.913	0.603	0.921	0.851	0.919
proposed model	**MSTGCN**	**0.895**	**0.854**	**0.843**	0.911	**0.645**	**0.924**	**0.866**	**0.924**

*The bold result is the best result and the underlined result is the second best result.