Enhancing Neurodegenerative Disease Diagnosis Through Confidence-Driven Dynamic Spatio-Temporal Convolutional Network

Ning Yuan, Donghai Guan, Shengrong Li, Li Zhang, and Qi Zhu, *Member, IEEE*

***Abstract*— Dynamic brain networks are more effective than static networks in characterizing the evolving patterns of brain functional connectivity, making them a more promising tool for diagnosing neurodegenerative diseases. However, existing classification methods for dynamic brain networks often rely on sliding windows to extract multi-window features, leading to suboptimal performance due to the spatio-temporal coupling on these windows and limited ability to effectively integrate complex topological features. To address these limitations, we propose a novel method called Confidence-Driven Dynamic Spatio-Temporal Convolutional Network (CD-DSTCN). First, our proposed method employs a spatio-temporal convolutional network integrated with a temporal attention mechanism to extract spatio-temporal features within each window. By propagating information across temporal windows during spatial convolution, the method effectively captures and integrates complex temporal and spatial dependencies. Second, each window generates an output probability, which quantifies prediction confidence based on the true class probability (TCP). This confidence score serves as a weight to assess the relative importance of different time windows. Finally, the confidence-weighted fused features are passed through a multilayer perceptron (MLP) for final classification. Extensive experiments on Alzheimer's and Parkinson's datasets show that the proposed method outperforms the state-of-the-art algorithms and can provide valuable biomarkers for brain disease diagnosis. Our code is publicly available at: https://github.com/YNingCode/CD-DSTCN**

Received 22 August 2024; revised 6 January 2025 and 11 March 2025; accepted 23 April 2025. Date of publication 28 April 2025; date of current version 7 May 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62472220, Grant 62371234, and Grant 62076129; in part by Jiangsu Province 100 Foreign Experts Introduction Plan under Grant BX2022012; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20231438; and in part by the Key Research and Development Plan of Jiangsu Province under Grant BE2022842. *(Corresponding authors: Donghai Guan; Qi Zhu.)*

Ning Yuan and Donghai Guan are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: dhguan@nuaa.edu.cn).

Shengrong Li and Qi Zhu are with the College of Artificial Intelligence and the Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: zhuqinuaa@163.com).

Li Zhang is with the Department of Geriatrics, Affiliated Brain Hospital of Nanjing Medical University, Nanjing 210029, China.

Digital Object Identifier 10.1109/TNSRE.2025.3564983

***Index Terms*— Dynamic brain network, spatio-temporal convolution, confidence fusion, spatio-temporal feature, neurodegenerative diseases diagnosis.**

I. Introduction

THE brain's high level of complexity can be attributed to its large number of neurons and intricate connections, forming a multi-layered, highly interconnected neural network [1]. This brain network can be viewed as a collection of nodes and edges. An in-depth understanding of how different brain regions are connected and their topology is crucial for exploring the brain mechanisms behind cognitive behaviors, and has gradually become an important tool for diagnosing and assessing brain diseases [2]. Recent studies have demonstrated that many neurodegenerative diseases and certain psychiatric conditions with complex causes, such as Alzheimer's disease, Amyotrophic lateral sclerosis and Parkinson's, are often associated with significant changes in brain network connectivity patterns [3], [4]. Therefore, it is important to investigate characterization methods that can accurately capture and differentiate the properties of brain networks.

Functional magnetic resonance imaging (fMRI) is a non-invasive brain imaging technique that monitors and records changes in blood oxygen levels during brain activity. Recently, it has been widely used to study the functional networks of the brain [5]. fMRI is based on the principle of blood oxygen level-dependent (BOLD) contrast, whereby different regions of the brain consume more oxygen during activity, leading to changes in blood oxygen levels in the corresponding regions. These changes can be captured and visualized by fMRI [6]. Recently, researchers have proposed a number of Static Functional Brain Network (SFBN) methods for the diagnosis of brain diseases [7], [8]. SFBN analysis methods focus on analyzing the functional connectivity of the brain in the resting state. In SFBN analysis, the brain is viewed as a whole network, which is usually measured based on the time-series correlation of BOLD signals. Such analyses help to reveal the intrinsic functional organization of the brain and potential disease markers. However, there is growing physiological evidence that brain functional networks

change over time, accompanied by rich temporal topological information that SFBNs cannot fully characterize [9], [10].

Therefore, in response to the properties of functional brain networks, researchers have begun to explore dynamic functional brain networks (DFBNs) [11]. Unlike SFBNs, DFBNs analysis focuses on temporal changes in brain functional connectivity, revealing a dynamically changing brain network structure over time. This provides a more detailed and comprehensive view to explore how the brain adapts to different task demands, state changes, or environmental stimuli [12]. For example, Jie et al. extracted temporal and spatial variability from each sub-network of the DFBN and integrated them into the identification of mild cognitive impairment (MCI) [13]. Noman et al. encoded the structure of each sub-network of the DFBN using a graph convolutional network (GCN), capturing the network topology's dynamic changes for autism spectrum disorder classification [14]. Most DFBN analysis methods are based on dividing the functional sequence signals into multiple time windows and constructing functional brain networks for the sub-sequences in each time window separately. However, brain activity occurs coherently, and the resulting network structure information is diffused along temporal and spatial dimensions throughout the brain. This means brain regions can have indirect effects on neighboring regions in the following time period through their interactions [11]. Therefore, focusing only on functional brain networks within a single time window cannot capture changes in the spatio-temporal topology of brain networks across time windows, nor can it effectively represent dynamic functional brain networks.

In general, existing works on feature extraction for dynamic brain networks can be divided into the following two categories. The first category involves extracting information directly from raw signal data, transforming it into feature vectors to represent brain information, as shown in Fig. 1(a). These approaches are difficult to effectively capture the deep spatio-temporal topological features in the complex information of dynamic brain networks [15]. The second category combines graph neural networks (GNNs), which have gradually emerged in recent years, to characterize the brain network as a graph structure, as shown in Fig. 1(b). These methods can capture deeper information features while preserving the topological information of the brain network [16], [17]. However, neither of these methods can effectively characterize the complex spatio-temporal topology of dynamic brain networks over time. Effectively learning and capturing the spatio-temporal topological features of dynamic brain networks remains a challenging task.

Facing this challenge, we design a confidence-driven dynamic spatio-temporal convolutional network for neurodegenerative disease diagnosis (CD-DSTCN), as shown in Fig. 1(c). Specifically, we first extract dynamic topological correlations from each window using a spatio-temporal graph network with a temporal attention mechanism [18]. Concurrently, in the spatial convolutional network, we develop a method for propagating information across windows, which can pass the features of the graph nodes in the spatial convolution of the current time window to the nodes of the next time window. This enables our network in the present to consider the effects of the brain network at previous moments, facilitating the extraction of long-lasting spatio-temporal information. In addition, we design a confidence fusion module capable of assigning a confidence score to each time window, indicating the level of importance of the features of each time window [19]. This module fuses the spatio-temporal topological features of each time window using the confidence scores. Finally the fused features are classified by a multi-layer perceptron. In brief, the method proposed in this paper has the following advantages:

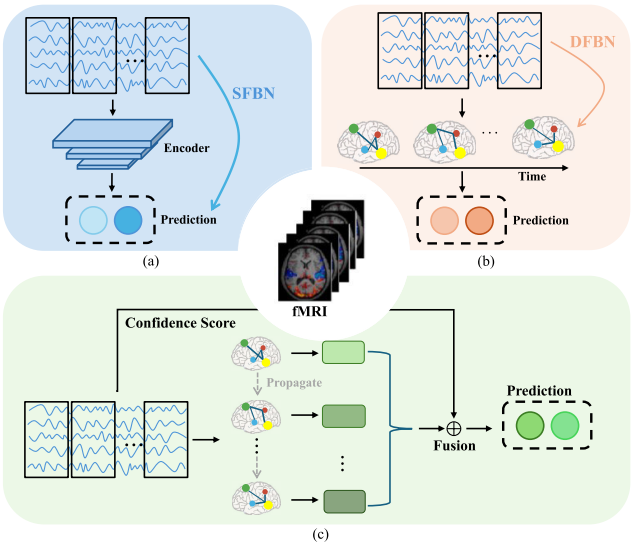


Fig. 1. Comparison of different brain network feature extraction strategies. (a) Static functional brain networks. (b) Constructing dynamic functional brain networks using graph structure. (c) The proposed dynamic spatio-temporal brain network model with confidence fusion propagating across time can extract complex spatio-temporal topological features of dynamic brain networks.

1) We propose a confidence-driven dynamic spatio-temporal convolutional network, which can learn the relationship between spatio-temporal topology over time in dynamic brain networks, effectively capturing the complex spatio-temporal topological features of these networks. These features have deep characterization capabilities and can be effectively applied to brain disease classification tasks.

2) Considering the connectivity of the brain network between time windows and constructing the dynamic brain network, the characterization of the spatio-temporal topological features extracted by the model is further improved through spatial convolutional information propagation across time windows.

3) We investigate a confidence fusion module that can efficiently fuse higher-order spatio-temporal topological features extracted from different time windows.

4) Our method has been evaluated for the diagnosis of cognitive dysfunction and Parkinson's disease. The results of multiple experiments have shown that our method is superior to current brain network analysis methods, identifying new spatio-temporal topological biomarkers in diagnosing brain diseases.

The remainder of this paper proceeds as follows: related work is described in Section II. Our proposed method is described in Section III. Section IV describes the dataset and experimental setup, and provides the experimental results for brain disease diagnosis. The discussions of the experimental results are given in Section V. Conclusions are presented in Section VI.

II. Related Works

A. Dynamic Brain Network

Dynamic brain networks refer to the time-varying patterns of functional connectivity between different regions of the brain. Unlike static brain networks, which assume a constant connectivity structure over time, dynamic brain networks account for fluctuations in connectivity that occur on timescales of seconds to minutes [20]. These fluctuations are believed to reflect the underlying neural processes that support cognitive functions and adapt to varying environmental demands [21]. Dynamic brain networks represent a promising avenue for enhancing the diagnosis and understanding of neurological and psychiatric disorders using fMRI. Research has shown that patients with Alzheimer's disease (AD) exhibit altered dynamic connectivity patterns compared to healthy controls. For instance, Jagust et al. demonstrated that AD patients have decreased variability in connectivity within the default mode network (DMN), which is crucial for memory and self-referential thoughts [22], [23]. In the diagnosis of schizophrenia, dynamic connectivity analysis has revealed that schizophrenia patients often show increased connectivity variability in networks associated with sensory processing and decreased variability in higher-order cognitive networks. Damaraju et al. found that these altered dynamic connectivity patterns could distinguish schizophrenia patients from healthy individuals with high accuracy [24].

Although several studies have successfully applied dynamic brain networks to the diagnosis of fMRI-related diseases [25], they have all focused solely on the spatio-temporal topology of the brain network within a single time window [26], ignoring the variation of the spatio-temporal topology between different time windows. In this paper, we propose spatio-temporal graph convolutional networks that propagate across time windows, fusing the spatio-temporal topological features of different time windows using confidence scores. This approach more effectively captures the complicated spatio-temporal topological features of dynamic brain networks.

B. Graph Convolutional Network

Graph Convolutional Networks (GCNs) are a type of neural network designed to operate on graph-structured data [27]. Traditional convolutional neural networks (CNNs) are highly effective for grid-structured data like images, where the spatial relationships between pixels are regular and consistent [28]. However, many real-world problems involve data with irregular structures, such as social networks [29], molecular structures [30], and brain networks [7], [8]. GCNs extend the concept of convolutions to graph data by applying convolution operations to the nodes of a graph, leveraging the graph's connectivity to learn node features.

GCNs can be divided into two categories: spatial-based graph convolution and spectral-based graph convolution. Spectral-based GCNs operate in the spectral domain of the graph, leveraging graph signal processing techniques to perform convolution operations. The key idea is to use the graph Laplacian, a matrix representation of the graph, to define convolution operations [31]. A spectral convolution on the graph signal x with a filter g_θ is defined as:

$$g_\theta * x = U g_\theta(\Lambda) U^T x \tag{1}$$

where U^T is the transpose matrix of eigenvectors of L, and Λ is the diagonal matrix of eigenvalues, $g_\theta(\Lambda)$ is a filter function applied to the eigenvalues. Spatial-based GCNs perform convolution operations directly on the graph's spatial structure [32]. These methods aggregate features from neighboring nodes to update the feature representation of each node, which can be expressed as follows:

$$Z = \sigma(\widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} X W) \tag{2}$$

where σ is an activation function, X is the input feature matrix, $\widetilde{A}$ is the adjacency matrix with added self-connections, $\widetilde{D}$ is the degree matrix, and W is the learnable weight matrix.

Due to the natural advantages of graph neural networks in dealing with non-Euclidean data and the complexity of brain networks at the spatial level, graph neural networks have been applied to brain network analysis in recent years. For example, Ktena et al. used a Siamese spectral domain GCN to extract the high-order feature by assessing the similarity between pairs of brain networks [33].

C. Uncertainty Estimation and Confidence

Confidence refers to the degree of certainty a model has regarding its predictions, which is a commonly used concept in deep learning applications, especially in the fusion of multimodal tasks. In the multimodal fusion model proposed by Choi and Lee [34], the confidence scores help the model distinguish which sensor data are closely related to the activity and which may be noisy or have little relevance to the current activity by evaluating the reliability of the preprocessed features. These scores effectively regulate the contribution of each sensor in the fusion process.

Confidence is also widely used in the diagnosis of neurological diseases and in neural signal classification tasks. Deng et al. introduced the Dirichlet distribution to characterize the expert category probability distribution in the diagnostic task of Frontotemporal Dementia (FTD), quantifying the confidence level of each view [35]. Song et al. used confidence levels to distinguish between trustworthy and untrustworthy samples in EEG signals for cross-center brain disease diagnosis [36].

The essence of all the aforementioned works is to use confidence levels to select modalities or samples with higher model trust for subsequent multimodal fusion or sample stratification. In our work, the amount and quality of information captured by different windows vary. By evaluating the confidence level

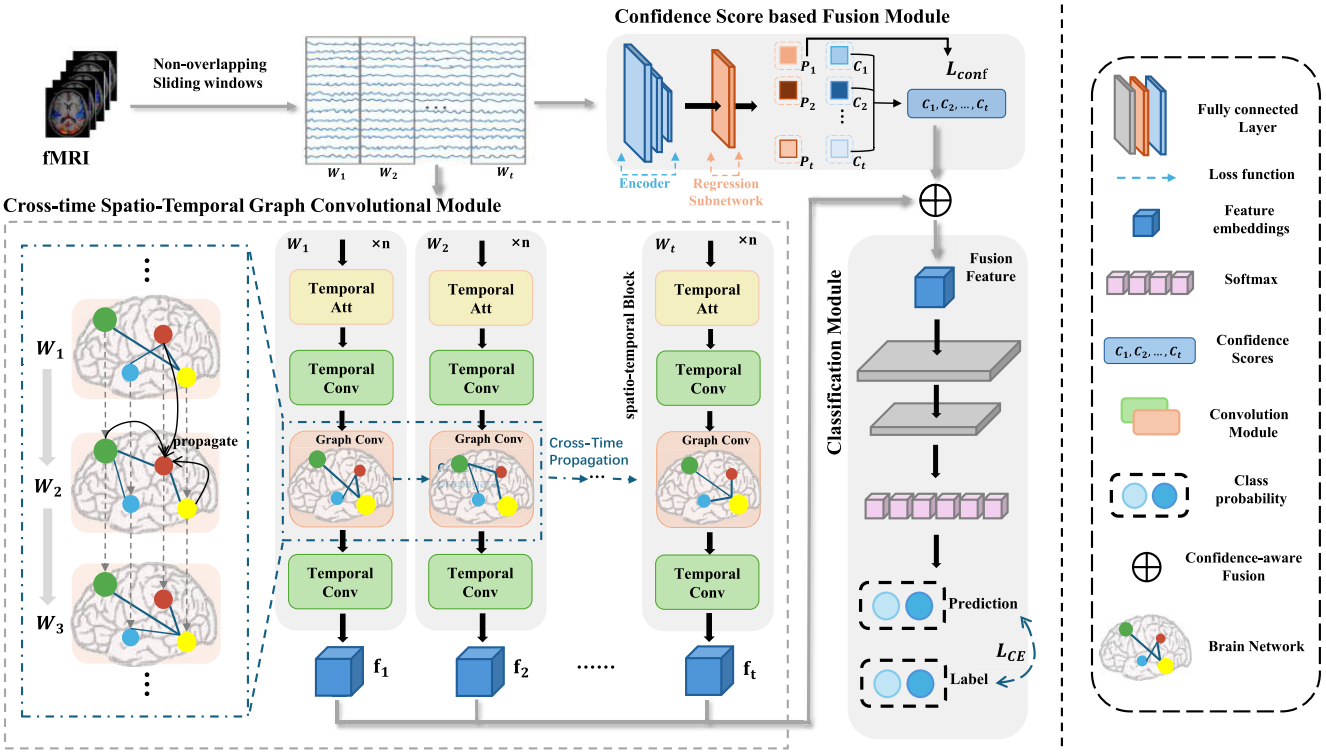


Fig. 2. The framework of the proposed confidence fusion for inter-temporal propagation of spatio-temporal convolutional dynamic brain networks.

of each window's features, these features can be weighted, allowing more reliable features to contribute more significantly to the final fusion result.

III. Proposed Method

In this paper, we propose a cross time spatio-temporal topology analysis method for dynamic functional brain networks. The framework of the proposed method is shown in Fig. 2. Our method can be roughly divided into a cross-time spatio-temporal graph convolution module and a confidence score based fusion module.

A. Cross-Time Spatio-Temporal Graph Convolutional Module

In this work, we assume that the rs-fMRI time-series data is $X = X_1, X_2, \ldots, X_N \in R^{N\times M}$, where N is the number of brain regions, and the M represents the length of time sequences. To characterize the dynamic changes in brain connectivity patterns, we first use T non-overlapping sliding windows of length L to divide the original rs-fMRI sequences.

1) Temporal Attention: The brain is a dynamic system that is constantly undergoing reconfiguration, with brain states changing over time and thus containing rich temporal information [13]. We design a temporal attention mechanism to capture important dynamic evolutionary information [37]. The temporal attention $Q \in R^{L\times L}$ (L means the timing length of each time window) is defined as follows:

$$Q = softmax(V_q \cdot sigmoid(XM_1)M_2(M_3X^T) + b_q) \quad (3)$$

where $V_q, b_q \in R^{L\times L}$, $M_1, M_2 \in R^N$, $M_2 \in R^{N\times N}$ denote the learnable parameters. We use the temporal attention matrix Q to capture important temporal information: $\hat{X} = (\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_T) = (X_1, X_2, \ldots, X_T)Q$.

2) Spatio-Temporal Convolution: This spatio-temporal convolutional network consists of temporal and spatial convolutions that capture information in the temporal and spatial dimensions, respectively [18]. The overall structure resembles a sandwich structure, with temporal convolution layer and spatial convolution layer.

Considering the temporal nature of dynamic brain networks, which are characterized by their capacity to process temporal information, the time-domain convolution part of the model employs 2D convolutional neural networks (2D CNNs) that are more adept at capturing contextual relationships [28]. For the spatial convolution component, we utilize a spatial domain GCN to gather spatial information.

For the sub-sequences of each time window, we adopt Pearson correlation coefficient [38] to construct functional brain network. Given the unique nature of each individual's brain functional connectivity patterns, we compute the adjacency matrix at the personal level. This approach helps network analysis and effectively captures the topological differences in brain networks across subjects. The Pearson correlation coefficient can be formulated as:

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X, \sigma_Y} = \frac{\mathrm{E}[(X - E_X)(Y - E_Y)]}{\sqrt{\sigma_X^2 \sigma_Y^2}} \quad (4)$$

where X and Y represent the data of two distinct brain regions, $cov(\cdot)$ represents covariance, σ represents standard deviation,

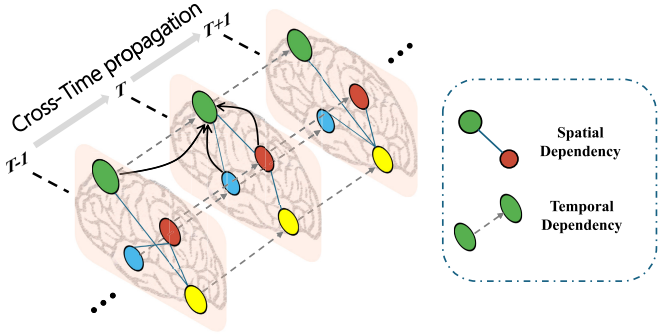


Fig. 3. Schematic flow of information propagation across time slices in spatio-temporal convolution. The solid blue line indicates the propagation of information from nodes within the same time slice, and the dashed black line indicates the propagation of information between the same nodes between neighbouring time slices.

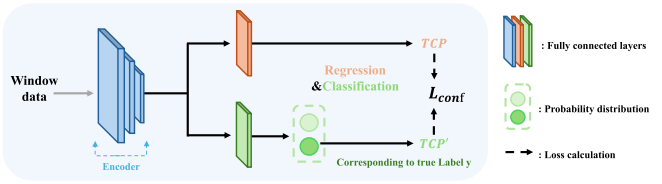


Fig. 4. The proposed Confidence-Score Module, which contains a pair of regression sub-network and classification sub-network. During the network training process, the classification sub-network is used to obtain the ground truth of TCP, and then the regression sub-network is trained to predict the TCP value by the loss function L_{conf}.

and E represents expectation. Thus, we can stack the brain networks across all windows to obtain dynamic functional brain networks: $P = (\rho_1, \rho_2, \ldots, \rho_T) \in R^{T\times N\times N}$. This matrix reflects the degree of correlation between the time series and can be used to construct a graph structure for subsequent spatial convolution.

To capture the evolutionary relationships of dynamic brain networks across time slices, we incorporated cross-time slice information propagation between the GCNs of different time slices, defined as follows:

$$Z^t = \widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}(X^t + Z^{t-1})\Theta \tag{5}$$

where X means the node feature matrix, which is derived from the output of the 2D CNN temporal convolution, $\widetilde{A}$ means the adjacency matrix with added self-connections, which is obtained from Pearson correlation and $\widetilde{D}$ is the degree matrix. The Z^{t-1} denotes the feature matrix that comes from the output of Spatial-GCN from the previous time window. The cross-time propagation process is shown in Fig. 3.

Finally, the spatio-temporal characteristics of each window can be expressed as:

$$f_t = CNN(ReLU(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}(CNN(X^t) + Z^{t-1})\Theta)) \tag{6}$$

where f_t means the output of the t^{th} window's spatio-temporal convolution, $ReLU(\cdot)$ means activation function.

B. Confidence Score Based Fusion Module

In order to more accurately fuse spatio-temporal topological features extracted from different time windows, we designed a confidence fusion module. This module evaluates the importance of spatio-temporal topological features from different time windows and assigns different weight values when fusing the features.

1) Confidence-Score Module: This module assesses the reliability of the data from each time window. In particular, a confidence score is assigned to each time window's prediction, reflecting the certainty of the prediction [19]. A higher confidence score indicates a lower degree of uncertainty in the features extracted from the corresponding window, thereby making the information from that window more reliable. The maximum class probability (MCP) can be employed to represent the confidence of the model's prediction as follows:

$$MCP(x) = \max_{x\in y} P(Y = k|w, x) = P(Y = \hat{y}|w, x) \tag{7}$$

where w denotes the network parameter set and $\hat{y}$ is the predicted class corresponding to sample x. Although MCP is effective for classification tasks, it may lead to overconfidence, especially for incorrect predictions. The use of the maximum SoftMax probability results in high confidence scores for both accurate and erroneous predictions. This indicates that the confidence criterion is unreliable. To address this issue, our method introduces true class probability (TCP) as a confidence criterion. In contrast to the MCP approach, which uses the maximum SoftMax output as the prediction confidence, the TCP method employs the output probability of the SoftMax corresponding to the true label as the prediction confidence. Formally, the true class probability (TCP) for each modality can be expressed as:

$$TCP^m(x^m, y) = P(Y = y|w, x^m) \tag{8}$$

where x^m denotes the high-dimensional feature vector of specific sample x, and y denotes its category.

When the classification results are correct, both TCP and MCP represent the maximum output of Softmax. However, when the classification is incorrect, TCP tends to approach a lower value, while MCP may assign a high confidence score to the incorrect classification. Although TCP provides more reliable confidence scores, it cannot be directly used during testing due to the absence of labeling information. To address this, we design a Confidence-Score Module, which consists of two MLP feature encoders and several fully connected layers, as shown in Fig. 4, to approximate the TCP. The two MLP encoders are used to predict the value of TCP and the probability distribution, respectively. We train this network using the MSE loss function:

$$L_{conf} = \frac{1}{T}\sum_{t=1}^{T}(c - c_{true})^2 \tag{9}$$

where c denotes the confidence-score predicted by the confidence regression network, c_{true} denotes the true value of the predicting confidence, and T means the number of windows.

We use the cross-entropy loss function for classification and, using the previously proposed loss function **Eq.(9)** as auxiliary

Algorithm 1 The Algorithm of the Proposed Method

Require: fMRI data, ground truth label Y, hyper parameters α, ϵ, the number of sliding window T, max_epoch;
Ensure: classification results and some evaluation metrics: ACC, PRE, RECALL, F1, AUC;

```
Divide the fMRI into T time slices {X_t}_1^T by non-overlapping sliding windows;
while epoch < max_epoch do
    Calculating the Pearson correlation coefficient matrix P = (p_1, p_2, ..., p_T) ∈ R^{T×N×N} by Eq.(4)
    for t=1 to N do
        Obtain valuable spatial information through the mechanism of temporal attention by Eq.(3)
        For each time window, extract spatio-temporal feature of {f_t}_1^T by Eq.(6)
    end for
    Each time slice obtains a Confidence Score {C_t}_1^T, and the L_conf through the CR Module by Eq.(9);
    Fuse features {f_t}_1^T based on the Confidence Score {C_t}_1^T by Eq.(11)
    Obtain the predicted label Y and calculate the L_CE;
    Calculate the final loss based on L_CE and L_conf by Eq.(10);
    Back-propagate final loss to update model parameters;
end while
return classification results and model parameters set;
```

supervision, we obtain the total loss function to be minimized:

$$L = L_{CE} + \alpha L_{conf} \tag{10}$$

where α is a hyperparameter, which adjusts the weight of L_{conf}.

2) Multi-Windows Features Fusion: We extract the spatio-temporal features of each window using the ST-Block, depicted as: $f = (f_1, f_2, \ldots, f_T) \in R^{N\times D}$. We then utilize a confidence regression network to dynamically estimate the prediction confidence for each window: $C_1, C_2, \ldots C_t$, which is used to guide the weighted fusion of different windows. This approach can enhance the predictions from the window with lower uncertainty by assigning them higher weights and suppresses those with higher uncertainty by assigning them lower weights, as follows:

$$f_{fusion} = \sum_{t=1}^{T} f_t C_t \tag{11}$$

where C_t represent the confidence of window t, f_t represent the feature of window t extracted form ST-Block, and T denotes the number of sliding windows.

Finally, several fully connected layers are used to learn the mapping relationship between the fused features and disease progression prediction. The overall framework of the method is shown in Algorithm 1.

C. Implementation

Our method is trained on a single GPU (NVIDIA GeForce RTX 3090), using a Python implementation based on Pytorch. The Adam algorithm is used for parameter optimization, and the rest of the relevant model training parameters are set as follows: the learning rate is 0.003, the weight decay is 3e-3, the epoch is 300, and the batch_size is 20. In the spatio-temporal convolutional network of the model, the temporal convolutional part uses a 2D CNN, with the convolutional kernel size set to (1, 3) and the step size set to (0, 1); the spatial convolutional part uses Spatial-based GCN.

For a more detailed overview of the specific parameter settings of the network, please refer to Appendix B. It contains comprehensive information on the proposed network architecture, including its components, parameter settings, and other configurations.

IV. Experiment

A. Materials

1) Datasets: In our work, we use the public Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (http://adni.loni.usc.edu/) and the Parkinson's disease (PD) dataset we collected for experimental validation. The specific details of the datasets are described below:

ADNI: The ADNI dataset is a widely used research resource designed to facilitate the early detection and monitoring of Alzheimer's disease (AD) and its related disorders. We use the data from the ADNI-2 and ADNI-3 phases of the dataset, containing data from a total of 203 samples, and used only fMRI data from the sample data. In the dataset, the number of normal controls (NC) is 73, and the number of patients is 130. The patients are divided into significant memory concern (SMC) with 60 patients and early mild cognitive impairment (EMCI) with 70 patients.

PD: The PD dataset used for the experiment is obtained from Affiliated Brain Hospital of Naniing Medical University. There are a total of 162 sample data in this dataset. A total of 52 individuals exhibited no abnormalities, they are identified as normal controls (NC), and the number of Tremor dominant Parkinson's disease (TDPD) patients is 44, the number of patients who exhibited postural instability and gait disorder (PIGDPD) characteristics is 64. The equipment used for data collection is a Siemens MAGNETOM Verio 3T scanner. The scanning parameters for rs-fMRI data are: repetition time = 2000 ms, echo time = 25 ms, and flip angle = 90°

2) Preprocessing: We perform the same preprocessing operations on the ADNI dataset and the PD dataset as described below: we use SPM9 from the DPARSF toolbox to preprocess all rs-fMRI data. The initial images are corrected and reorganized by dividing the serially stored data into distinct sections and applying adjustments in accordance with the EPI template. The effects of head motion and cerebrospinal fluid and white matter disturbances are reduced by deskew processing. Then, we use AAL mapping to divide the rs-fMRI of ADNI into 90 ROIs, each containing 197 time points, and the rs-fMRI of PD into 116 ROIs, each containing 220 time points.

B. Methods for Comparison

To validate the effectiveness of our proposed method, we compare it with several brain network feature extraction

TABLE I
THE CLASSIFICATION RESULTS OF THE PROPOSED METHOD AND SOTA METHODS ON THE ADNI DATASET (%)

	Methods	NC vs. SMC & EMCI			NC vs. SMC			NC vs. EMCI			SMC vs. EMCI		
		ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
SFBN	SLR	75.86	66.21	81.59	74.83	74.65	82.13	78.95	78.13	85.98	65.38	64.00	70.60
	PC	77.00	67.24	82.76	76.64	76.56	83.63	71.73	69.46	79.31	55.85	51.18	59.10
	FBNetGen	84.33	88.74	79.70	78.76	74.98	77.97	84.89	76.96	86.21	66.92	53.12	60.36
	GraphSAGE	85.66	88.49	86.20	86.71	83.93	88.21	89.63	**87.81**	**90.41**	71.53	60.96	69.13
	SAGPool	85.21	88.65	88.07	85.33	81.01	86.43	87.85	76.34	81.24	70.00	63.49	62.01
	tHOFC	77.39	68.09	81.74	76.71	76.79	82.66	73.23	69.79	78.53	56.46	49.88	54.30
	aHOFC	75.67	64.69	81.02	81.12	80.75	85.54	71.58	68.69	78.34	56.15	52.35	57.40
	GCN	81.66	84.12	85.75	81.71	75.88	78.84	78.90	68.80	79.76	70.15	63.44	66.84
	GAT	86.78	87.12	87.74	83.90	81.12	84.39	86.21	87.31	83.47	72.30	63.54	64.40
DFBN	DFC	75.28	80.68	82.52	82.92	80.17	84.58	70.66	64.48	82.54	56.95	60.28	63.79
	CGRL	82.65	81.27	83.58	86.74	86.23	84.20	84.86	79.90	84.26	74.43	71.33	74.35
	FE-STGNN	82.46	87.53	85.65	81.25	84.32	85.24	82.02	80.00	81.48	74.56	72.00	73.67
	ST-fMRI	87.15	89.36	86.98	86.27	85.09	87.45	88.04	83.45	88.72	76.02	73.28	72.23
	DHOFC	71.92	62.25	77.11	72.03	70.59	75.75	69.92	66.67	72.88	53.85	46.43	56.83
	ST-GCN	85.83	87.42	83.57	88.14	85.98	85.16	84.06	82.23	84.50	72.46	69.38	72.76
	Ours	**88.21**	**89.86**	**88.96**	**89.64**	**86.87**	**88.79**	**89.85**	87.68	89.82	**78.23**	**75.00**	**76.46**

TABLE II
THE CLASSIFICATION RESULTS OF THE PROPOSED METHOD AND SOTA METHODS ON THE PD DATASET (%)

	Methods	NC vs. TDPD & PIGDPD			NC vs. TDPD			NC vs. PIGDPD			TDPD vs. PIGDPD		
		ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
SFBN	SLR	72.84	61.40	77.81	62.71	62.07	74.83	64.29	58.82	76.35	54.63	42.35	59.02
	PC	67.35	52.51	71.43	69.49	71.08	76.72	63.88	61.18	67.09	59.44	48.17	60.83
	FBNetGen	80.84	80.17	75.27	74.39	72.50	69.74	78.67	65.99	71.56	75.90	60.53	61.21
	GraphSAGE	77.35	80.24	68.85	77.87	75.69	74.41	77.77	74.07	72.87	74.09	60.18	64.69
	SAGPool	76.54	82.11	71.56	73.71	71.24	75.17	78.55	71.53	78.09	76.90	59.63	65.78
	tHOFC	68.02	54.66	71.54	50.08	59.02	62.98	66.22	62.49	71.90	61.57	62.10	65.50
	aHOFC	66.73	53.06	73.89	67.71	61.62	71.52	67.86	65.03	73.61	60.07	61.39	60.79
	GCN	78.41	76.27	76.73	79.19	79.61	77.42	78.66	63.01	76.57	71.36	59.41	66.00
	GAT	80.50	82.80	75.24	77.95	69.48	73.42	78.77	75.22	77.32	74.63	56.34	66.58
DFBN	DFC	70.35	68.46	71.27	68.72	56.26	60.73	68.45	66.78	69.21	58.90	56.06	58.43
	CGRL	76.31	74.22	74.17	74.06	71.39	72.36	71.90	70.13	67.16	63.43	58.71	60.25
	FE-STGNN	73.86	79.28	66.17	72.80	70.25	68.06	75.73	65.07	71.90	68.77	64.25	55.56
	ST-fMRI	76.12	78.06	74.28	72.79	67.33	69.98	73.84	67.06	71.79	75.32	64.33	**71.32**
	DHOFC	62.47	52.15	71.03	59.68	61.67	72.34	61.03	59.76	68.32	55.20	60.73	61.45
	ST-GCN	80.11	82.65	75.24	76.89	78.56	77.49	80.88	74.12	77.80	73.36	50.25	56.24
	Ours	**81.20**	**83.55**	**80.32**	**82.71**	**83.76**	**80.99**	**84.77**	**76.72**	**79.38**	**77.00**	**65.61**	69.95

methods. Contains 9 static brain network feature extraction methods: SLR [39], PC [38], FBNetGen [40], GraphSAGE [32], SAGPool [41], tHOFC [42], aHOFC [43], GCN [31], GAT [44], and 6 Dynamic brain network feature extraction method: DFC [45], CGRL [46], FE-STGNN [47], ST-fMRI [25], DHOFC [48], ST-GCN [49]. Among them, SLR, PC, tHOFC, aHOFC, and DHOFC use the Brain Network Construction & Classification toolbox [50].

C. Experimental Setup

For all methods, we use a 10-fold cross-validation approach, where 8-folds is used as the training set, 1-fold as the validation set, and the remaining 1-fold as the test set. Specifically, to ensure consistency, we use the Pearson correlation coefficient method to obtain the adjacency matrix in the parts where the graph structure is constructed.

On the ADNI dataset, we conduct four classification tasks: NC vs. SMC & EMCI, NC vs. SMC, NC vs. EMCI and SMC vs. EMCI. On the PD dataset, we perform four classification tasks: NC vs. TDPD & PIGDPD, NC vs. TDPD, NC vs. PIGDPD and TDPD vs. PIGDPD. To evaluate the classification performance of the model, three evaluation metrics: Accuracy (ACC), F-Measure (F1), and the Area Under the Receiver Operation Characteristic Curve (AUC), are used to assess the model performance. In Appendix C, we supplement the experimental results of macro-level evaluation metrics. Higher values of these metrics indicate better model performance.

D. Classification Performance

Table. I and Table. II list the classification results of this paper's method with the other 16 comparative methods on the two datasets ADNI and PD. As can be seen from the tables, the classification accuracies of our method on the ADNI dataset are 88.21%, 89.64%, 89.85%, and 78.23% for the four tasks, respectively. For the PD dataset, the classification accuracies are 81.20%, 82.71%, 84.77%, and 77.00%. Compared to all other methods, our proposed method achieves the highest accuracy. Additionally, we plot the classification scatter plots of the spatio-temporal topological features extracted by our method. As seen in Fig. 5, the scatterplot of ST-GCN shows some classification effect, but the feature points of different classes are more dispersed and do not aggregate well. The GAT method shows better aggregation of feature points but does not maintain strong classification performance. In contrast, the spatio-temporal topological features extracted by our method

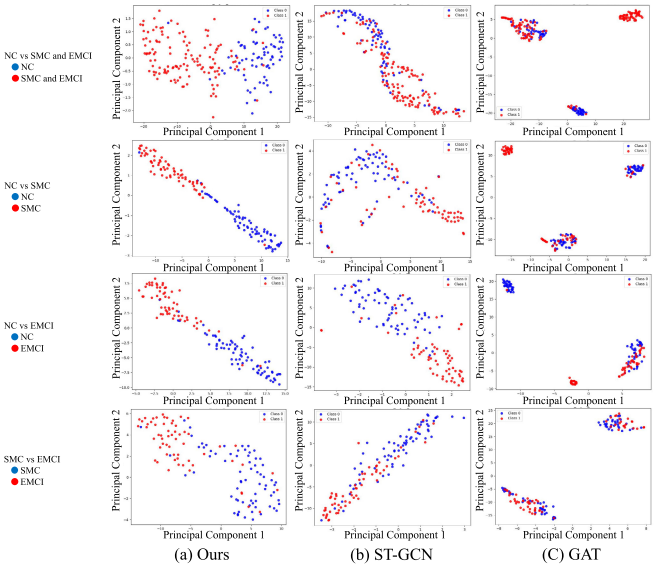


Fig. 5. Scatterplot of spatio-temporal features extracted by our method versus other methods for four classification tasks on the ADNI dataset. In order from top to bottom: NC vs. SMC & EMCI, NC vs. SMC, NC vs. EMCI and SMC vs. EMCI.

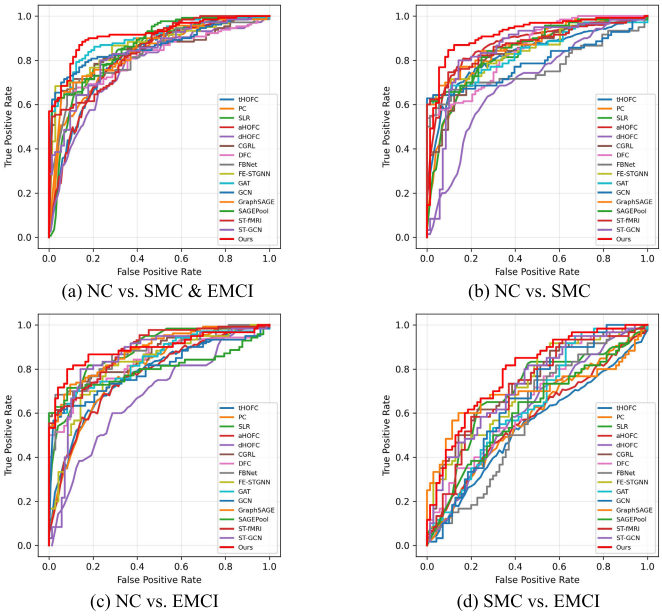


Fig. 6. ROC curves of the proposed method and comparison algorithms for different tasks on the ADNI dataset.

demonstrate excellent classification performance, effectively clustering similar features and revealing an intrinsic pattern or structure between classes.

On this basis, we plot the ROC curves of our method and all other comparative methods for the four classification tasks on the ADNI dataset. The results are shown in Fig. 6. As seen in Fig. 6, the ROC curve of our method is positioned in the upper left corner among all comparative methods, indicating the highest AUC value. This demonstrates the superior classification performance of our proposed method in identifying Alzheimer's disease and its subtypes.

The experimental results demonstrate that our method achieves excellent outcomes in identifying Alzheimer's and Parkinson's diseases and their subtypes. This improvement is attributed to our method's ability to capture the spatio-temporal dynamic evolutionary information from the dynamic brain network. To preserve the long and short temporal information that might be lost due to the piecewise construction of the brain network, we propagate topological information across time slices. This allows the spatio-temporal convolutional network to account for information from intermediate time slices. Additionally, when constructing the brain network in slices, we design a confidence module to focus on the more important time slice, thereby obtaining more significant spatio-temporal topological information. Within each time slice, the temporal attention mechanism helps the model autonomously focus on the critical spatio-temporal information. Consequently, our model can extract more discriminative spatio-temporal topological features more efficiently, leading to superior classification performance.

To assess the generalizability of our method to other neural data modalities, we conduct supplementary experiments on a motor imagery (MI) EEG dataset. The detailed experimental setup, results, and analysis for the EEG data are presented in Appendix. A.

E. Ablation Studies

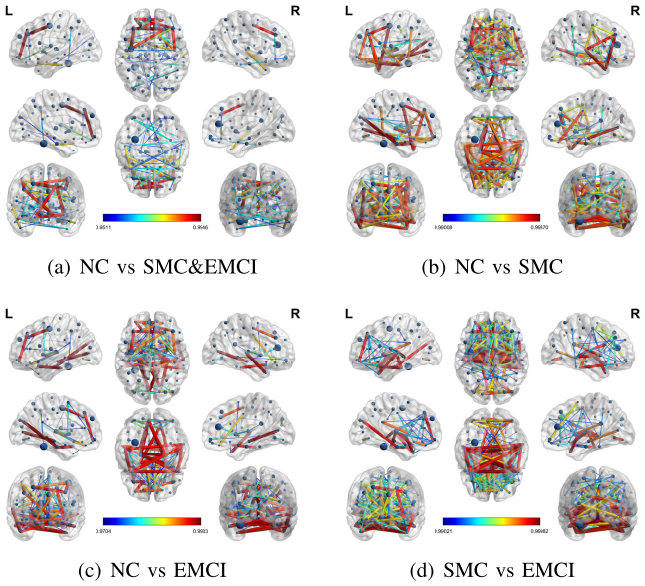
To verify the effectiveness of the confidence module, topological information propagation across time slices, and the temporal attention mechanism in our proposed method, we conduct ablation experiments on both ADNI and PD datasets. The experimental results are shown in Table. III. In the first variant "w/o Confidence Score", we exclude the use of the confidence module and use simple concatenation when fusing temporal topological features extracted from different time slices. In the second variant "w/o Cross-time propagation", we exclude the propagation of information across time slices in the spatio-temporal convolution and simply use the spatio-temporal convolution network to extract features in each time slice separately. In the third variant "w/o Temporal attention", we exclude the use of temporal attention mechanisms in the spatio-temporal convolutional networks. In the subsequent variants, we exclude multiple modular mechanisms to further investigate their individual and combined contributions.

From Table. III, it can be seen that in the four classification tasks on both ADNI and PD datasets, the performance of the model variants decreased by 6%-8% compared to the full model after the exclusion of the confidence module. This demonstrates that the confidence module can efficiently fuse the spatio-temporal topological features from different time slices, better accounting for the importance of the spatio-temporal topological information of the brain network compared to simple concatenation. In addition, the performance of the model variants decreases by 1%-3% after excluding the propagation of information across time slices. This proves that the propagation of information across time slices enables the current spatio-temporal convolution module to effectively consider the spatio-temporal topological information of the previous time slices, leading to a better

TABLE III
THE ACCURACIES OF ABLATED MODELS ON THE TWO DATASETS (%)

Dataset	Method	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
		NC vs. SMC & EMCI			NC vs. SMC			NC vs. EMCI			SMC vs. EMCI		
ADNI	w/o Confidence Score	82.71	83.66	83.26	82.42	74.93	78.08	84.06	78.08	76.90	70.00	54.39	64.02
	w/o Cross-time propagation	87.73	87.33	87.59	88.42	85.19	86.54	86.89	81.31	85.32	78.46	73.51	75.78
	w/o Tempotal attention	87.26	87.56	87.10	87.42	85.17	87.19	88.08	82.59	87.79	76.53	72.56	75.54
	w/o Con&Cross-p	84.61	83.36	85.27	83.66	83.95	85.05	84.50	83.70	84.61	74.15	60.47	70.58
	w/o Con&Tem	81.28	84.62	80.47	82.38	79.34	82.49	79.78	70.67	76.53	74.61	68.27	59.82
	w/o Cross-p&Tem	85.16	87.82	85.09	86.99	83.53	84.79	83.46	76.74	80.01	75.15	71.40	71.12
	Ours	**88.21**	**89.86**	**88.96**	**89.64**	**86.87**	**88.79**	**89.85**	**87.68**	**89.82**	**78.23**	**75.00**	**76.46**
		NC vs. TDPD & PIGDPD			NC vs. TDPD			NC vs. PIGDPD			TDPD vs. PIGDPD		
PD	w/o Confidence Score	75.39	77.29	70.00	74.09	77.38	68.81	76.28	71.50	75.36	74.02	71.06	69.77
	w/o Cross-time propagation	80.39	81.59	78.92	81.21	81.08	75.27	80.55	74.93	77.11	75.81	62.33	62.67
	w/o Tempotal attention	80.58	79.52	70.86	81.58	80.52	73.86	80.88	73.06	78.33	72.18	62.56	**70.54**
	w/o Con&Cross-p	78.63	81.64	66.23	78.86	76.78	74.21	76.77	75.61	74.72	71.63	60.20	65.94
	w/o Con&Tem	75.99	78.41	60.27	72,87	68.00	67.74	73.66	50.24	64.24	73.00	49.04	63.81
	w/o Cross-p&Tem	76.77	79.72	73.05	72.27	61.58	64.95	76.88	65,30	74.67	73.90	58.75	63.89
	Ours	**81.20**	**83.55**	**80.32**	**82.71**	**83.76**	**80.99**	**84.77**	**76.72**	**79.38**	**77.00**	**65.61**	69.95

* "w/o" means "without"



(a) NC vs SMC&EMCI (b) NC vs SMC

(c) NC vs EMCI (d) SMC vs EMCI

Fig. 7. Brain network construction on the ADNI dataset's different tasks.

characterization of the extracted features. After removing the temporal attention mechanism, the model performance decreases by 1%-3%, highlighting the necessity of capturing important spatio-temporal information within each time slice. Further experimentation reveals that removing any two of these three components results in varying degrees of decrease in model classification accuracy.

These results highlight the importance of the three methods in our approach to modelling dynamic brain networks and capturing their spatio-temporal topological features.

V. DISCUSSION

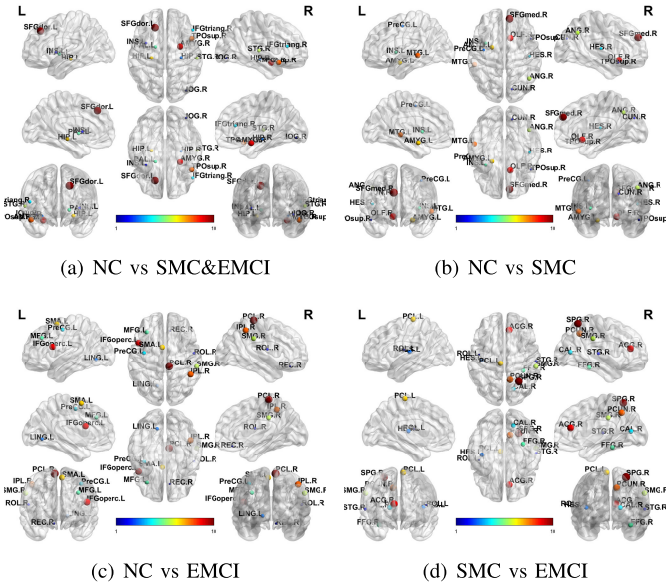
A. Visualization of Dynamic Brain Networks

To more intuitively demonstrate the construction of dynamic brain networks and the changes in their spatio-temporal topology, we visualize the brain networks constructed for the four disease classification tasks in the ADNI dataset, as shown in Fig. 7. From Fig. 7, it is evident that the connection weights between different brain regions vary, with tighter connectivity observed between brain regions that exhibit stronger correlations and weaker connectivity between regions with lesser correlations. This visualization reflects, to some extent, the complex connectivity patterns between real brain lobes.

As shown in Fig. 7, the connectivity constructs of the dynamic brain network for the ADNI disease classification task reveal that connections with higher weights are mostly concentrated in the prefrontal lobe. This is due to the central role of the prefrontal lobe in cognitive control, executive function, and emotion regulation, as well as the significant impact of pathological changes in Alzheimer's disease on prefrontal lobe [51]. Abnormal changes in prefrontal connectivity not only reflect local dysfunction but also reveal broader effects on the functioning of whole brain networks, providing an important neurobiological basis for understanding and diagnosing Alzheimer's disease. Additionally, we observe that connections with higher weights are not only confined within particular brain lobes but also occur between different lobes, such as regions between the temporal and occipital lobes and regions between the left and right temporal lobes. This is because the temporal lobes, especially the hippocampal region, are closely associated with memory formation and retrieval. In Alzheimer's disease, there is impaired integration of information across multiple brain regions, leading to increased connectivity between the frontal, temporal, and occipital lobes. This pattern may represent the brain's compensatory mechanisms, attempting to maintain coordination between different cognitive functions by enhancing communication across these regions.

B. Discriminative Brain Regions

To effectively assess the importance of each node in the brain network, we first extract brain network features using our method for the four disease classification tasks on the ADNI dataset. We then apply the PageRank algorithm to identify the top 10 most discriminative brain region nodes and visualized these nodes, as shown in Fig. 8. Identifying these important nodes is significant for understanding the complex network structure of brain function, particularly the



(a) NC vs SMC&EMCI (b) NC vs SMC

(c) NC vs EMCI (d) SMC vs EMCI

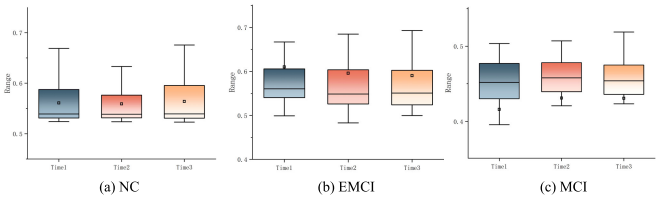
Fig. 8. The 10 most important brain region nodes for different classification tasks on the ADNI dataset.

network abnormalities in Alzheimer's disease patients. This visualization not only highlights the status of these nodes but also provides targets for further functional analysis and intervention.

As seen in Fig. 8, the nodes of higher importance in the ADNI classification task are concentrated in the prefrontal and temporal lobes. This finding aligns with existing scientific literature [52]. The prefrontal lobe play a key role in higher cognitive functions, including decision-making, working memory, attentional control, and executive functions. Additionally, the prefrontal lobe is typically centrally located in brain networks with extensive connectivity, forming strong functional connections with multiple other brain lobes such as the parietal lobe, temporal lobe, and limbic system [53]. The critical role of the temporal lobe in integrating and storing information makes it particularly important in Alzheimer's disease (AD), as early symptoms of AD often include memory impairment. Connections between the temporal lobe and prefrontal lobe also play an important role in higher cognitive functions, leading to the high importance of nodes in these lobes.

In Fig. 8, high importance nodes are also observed around the parietal lobe and Wernicke's area. The parietal lobe plays a crucial role in spatial cognition, body coordination, and visual information integration. Alzheimer's disease (AD) patients often experience a decline in spatial perception and navigation, which is closely associated with impaired function of the parietal lobe. Wernicke's area, located in the posterior part of the left superior temporal gyrus, is primarily responsible for language comprehension and processing. Language deterioration in AD patients is often accompanied by lesions in Wernicke's area, manifesting as difficulties in vocabulary retrieval and semantic confusion.

Our method successfully extracts brain network features, identifying nodes of higher importance in the frontal, temporal, and parietal lobes. This demonstrates the method's capability to accurately capture real brain functional connectivity and



(a) NC (b) EMCI (c) MCI

Fig. 9. Confidence scores for different time slices in different tasks on the ADNI dataset.

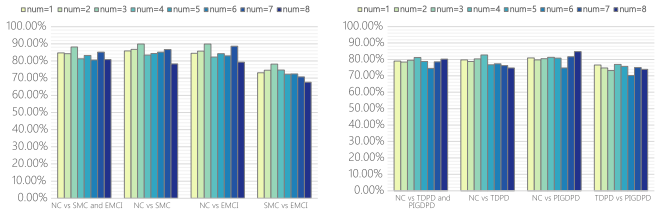
effectively identify disease-related nodes. Additionally, these high-importance brain region nodes, which are often central within complex neural networks, exhibit extensive connectivity with multiple other functional regions. This connectivity underscores the critical role these nodes play in information dissemination and integration across brain-wide networks. Therefore, the prominent ranking of brain regions near the frontal, temporal, parietal, and Wernicke's areas reflects not only their individual functional importance but also their pivotal role in the integration of information across regions.

C. Confidence-Scores of Different Time-Slices

In this study, we assign confidence scores to each time segment for the classification tasks in the ADNI dataset using the methodology proposed in this paper. These confidence scores are visualized through box plots, and the results are shown in Fig. 9.

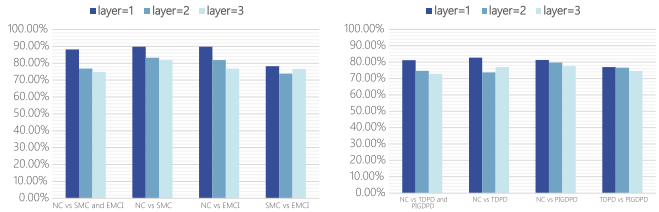
As seen in Fig. 9, there are significant differences in the confidence scores of different classification groups across the time segments. In the normal control (NC) group, the confidence scores of the three time segments were approximately the same, indicating that the brain functional connectivity of normal individuals remained stable over time with no obvious fluctuations in functional activities, reflecting a high degree of temporal consistency. For early mild cognitive impairment (EMCI), the confidence score for time segment 1 was higher than the other two segments. This suggests that the network in this paper focused more on functional changes in the early stages when analyzing the EMCI group, possibly because EMCI patients exhibit specific brain functional features early in cognitive tasks, which help distinguish them from other groups [54]. In the significant memory concern (SMC) group, the confidence score for time segment 3 was higher than the first two segments, indicating that our network focused more on functional changes in the later stages when analyzing the SMC group. This may reflect that SMC patients show more functional brain activity or compensatory activity in the final stages of cognitive tasks [55].

Our method not only demonstrates the dynamic changes in brain functional connectivity across different time segments by analyzing the confidence scores but also highlights the superiority and usefulness of our network in classification tasks. This approach offers new perspectives for understanding brain functional characteristics in various cognitive states and provides robust support for the early diagnosis and intervention of Alzheimer's disease.



(a) accuracies on ADNI dataset (b) accuracies on PD dataset

Fig. 10. The accuracies (%) of our method with different numbers of windows_num under the four diagnosis tasks.



(a) accuracies on ADNI dataset (b) accuracies on PD dataset

Fig. 11. The accuracies (%) of our method with different numbers of ST-Blocks under the four diagnosis tasks.

D. Sensitivity Analysis

In this subsection, we perform a sensitivity analysis on the parameters in the model: the number of windows num and the number of ST-Block layers. The experimental results of this sensitivity analysis can be seen in Fig. 10 and Fig. 11.

1) Number of Windows Num: To capture the dynamic nature of brain networks, we divide the fMRI data with non-overlapping sliding windows to construct dynamic brain networks. To evaluate the effect of different numbers of windows on the model classification performance, we consider the window sizes from 1 to 8, the results in Fig. 10 indicate the optimal performances are obtained when the window number between 2 to 4 on the ADNI and PD datasets. While the choice of window number affects the classification, the effects is relative small (within 5%). This is because when the number of windows is small, there are more time points in each window, making it difficult for the spatio-temporal convolutional network to capture long-term spatio-temporal information. Conversely, when the number of windows is high, there are fewer time nodes in each window, but too many windows can spread the data too thinly, losing the continuous spatio-temporal information in the original data.

2) Number of ST-Block Layers: In our approach, the spatio-temporal convolutional network for information propagation across time slices can use multiple layers. To evaluate the effect of different numbers of ST-Block layers on classification performance, we set the number of layers to [1, 2, 3], and the experimental results are shown in Fig. 11. Results indicate that a single ST-Block consistently achieves the best performance. This suggests that increasing network depth does not necessarily lead to improved results in this context. Our model incorporate a spatial graph convolution network (GCN) component. As the number of spatio-temporal convolutional network layers grows, the spational convolution becomes prone to overfitting. Consequently, the information of each network node contains excessive information about its neighbors and N-hopping neighbors, causing the extracted spatio-temporal topological features to lose their good characterization ability.

VI. Conclusion

To effectively capture the spatio-temporal topological information in fMRI dynamic brain networks, we develop a novel confidence fusion-based spatio-temporal convolutional network model that propagates across time windows, for the diagnostic task of neurodegenerative diseases. Our method involves dividing fMRI data into non-overlapping segments, constructing dynamic brain networks for each segment, and extracting spatio-temporal topological features using a spatio-temporal convolutional network with temporal attention. To enhance feature representation, we incorporate the propagation of spatio-temporal information across time slices, allowing the network to consider historical context. Additionally, the spatio-temporal features of each time window are assigned different confidence levels to weight their contribution to the final representation by the confidence module. Experimental results on the ADNI and PD datasets demonstrate that our method outperforms state-of-the-art methods in identifying and diagnosing Alzheimer's and Parkinson's diseases, highlighting its effectiveness in extracting and characterizing spatio-temporal topological features of dynamic brain networks.

Appendix

A. Supplementary Experiments

To further validate the generalizability of the proposed method beyond fMRI data, we conduct experiments on an additional EEG dataset. To ensure consistency in the experimental setup, we choose not to use the widely adopted differential entropy (DE) feature, commonly utilized in EEG classification tasks. Instead, we directly employ the raw processed signal data.

1) Datasets: We use the publicly available motor imagery EEG datasets from Shanghai University: shu_dataset [56]. The dataset comprises 25 subjects, each trains five times over a span of 2-3 days between sessions. Each session includes 100 trials, focusing on motor imagery tasks involving left-hand (L) and right-hand (R) movements. EEG datas are recorded using 32 electrodes based on the International 10-10 system, with a sampling rate of 250 Hz and a 4-second time window per trial. Each subject completes a total of 500 trials (5 sessions $\times$ 100 trials), with each trial consisting of 1,000 sampling points.

2) Experimental Setup: We perform within-subject experiments on the EEG dataset, calculating classification accuracies for each subject individually. The final performance metric is derived by averaging the accuracies across all 25 subjects.

The experimental results are presented in A.Table. IV. As shown in the table, our method outperforms other SOTA methods designed for fMRI, achieving the highest average

TABLE IV
THE CLASSIFICATION RESULTS OF THE PROPOSED METHOD AND SOTA METHODS ON THE EEG DATASET (SHU_DATASET) (%)

Subject	Method															
	SFBN									DFBN						
	SLR	PC	FBNetGen	GraphSAGE	SAGPool	tHOFC	aHOFC	GCN	GAT	DFC	CGRL	FE-STGNN	ST-fMRI	DHOFC	ST-GCN	Ours
Sub1	46.75	56.95	68.93	61.13	61.17	58.07	62.64	64.02	60.55	54.72	61.02	67.28	67.13	62.18	65.24	**69.30**
Sub2	57.36	61.96	75.73	59.12	60.79	53.70	54.12	69.08	63.70	58.37	65.43	72.93	69.71	72.41	**76.34**	74.11
Sub3	54.75	58.82	58.54	58.26	59.68	47.65	47.83	66.31	60.74	58.33	63.68	63.11	65.88	58.56	61.76	**67.06**
Sub4	48.26	51.63	60.79	59.34	60.84	49.45	48.37	65.29	60.63	60.29	64.12	66.19	65.65	61.42	67.79	**68.07**
Sub5	52.64	51.26	67.42	62.12	64.22	50.36	50.13	**73.05**	59.33	61.22	68.25	69.58	70.95	53.26	71.38	72.31
Sub6	58.86	52.75	87.11	61.84	64.82	51.33	50.27	90.41	86.55	58.98	71.84	86.56	86.56	56.75	**91.87**	89.41
Sub7	56.10	67.98	63.52	60.45	62.89	69.89	68.82	71.52	63.94	60.55	66.95	62.83	70.12	67.79	**73.36**	70.21
Sub8	54.61	57.05	68.48	60.12	60.51	58.35	48.39	69.29	58.04	56.88	62.19	69.07	68.73	65.95	69.10	**71.19**
Sub9	44.80	47.40	58.01	59.42	64.08	46.93	49.75	64.69	57.82	58.21	57.11	61.72	62.41	58.42	60.46	**68.61**
Sub10	45.12	52.71	60.54	61.87	61.66	58.25	55.48	62.29	61.66	55.76	58.03	60.61	65.29	62.50	**67.50**	60.41
Sub11	47.75	48.34	58.26	59.12	62.95	47.55	45.79	66.39	59.92	62.01	64.45	60.03	65.84	58.38	65.74	**68.73**
Sub12	53.75	50.19	61.35	58.73	60.91	51.13	49.84	63.32	60.91	61.13	62.84	68.42	67.49	63.39	65.27	**70.36**
Sub13	57.37	65.34	**85.69**	62.35	66.24	57.32	58.36	81.78	75.65	53.92	75.77	74.26	78.65	73.14	83.02	84.67
Sub14	47.56	56.09	60.05	61.38	63.26	53.42	51.32	68.69	64.50	57.75	65.56	61.92	70.11	53.69	61.80	**72.41**
Sub15	54.22	51.24	70.17	60.05	64.26	51.20	50.60	**72.27**	64.05	59.47	66.29	68.54	68.96	52.84	67.85	70.41
Sub16	47.03	62.37	71.00	61.52	62.34	57.10	57.29	70.16	64.59	60.91	68.56	70.72	71.47	71.49	67.47	**72.70**
Sub17	52.88	47.15	59.60	61.63	64.07	49.83	46.50	65.69	60.22	61.35	63.15	64.98	64.59	59.78	64.71	**69.21**
Sub18	54.01	64.77	67.42	59.88	61.94	63.22	59.55	**70.65**	67.19	56.22	62.49	70.41	69.84	61.68	65.79	70.31
Sub19	52.96	70.32	72.28	61.85	63.53	66.82	62.47	75.48	69.31	58.61	67.82	71.13	70.36	67.26	74.20	**77.26**
Sub20	49.69	82.31	75.08	81.19	63.39	81.06	81.68	87.17	88.13	59.04	74.71	67.88	78.77	73.94	85.47	**88.41**
Sub21	50.34	58.47	77.13	61.96	63.70	69.12	70.60	81.93	75.93	56.87	73.51	71.54	78.21	65.07	80.11	**84.18**
Sub22	49.29	51.58	71.04	59.66	62.08	49.18	51.58	75.41	71.25	59.32	66.18	72.47	71.02	55.58	73.95	**77.91**
Sub23	53.23	52.86	**83.57**	65.16	62.55	59.35	52.90	81.23	76.53	56.15	63.24	71.24	70.23	66.91	74.91	75.02
Sub24	45.16	47.65	59.87	61.96	64.05	47.08	46.78	64.99	61.56	62.34	58.22	66.99	67.42	56.75	65.66	**69.06**
Sub25	53.43	49.45	60.94	58.45	62.99	50.01	50.51	**67.55**	60.82	58.75	62.85	62.64	65.32	63.21	61.25	60.50
Avg	51.51	56.66	68.10	61.54	62.75	55.89	54.86	71.54	66.14	58.68	65.37	68.12	70.02	62.49	70.48	**72.87**

TABLE V
THE FRAMEWORK DETAILS OF CD-DSTCN ON ADNI DATASET

Module	Layer	Kernel Size	Output (B, T/N, D)
Confidence Score based Fusion Module	Uncertainty Layer		[20, 3, 1]
	Classification Layer		[20, 3, 2]
Cross-time Spatio-Temporal Graph Convolutional Module	Temporal Attention		[20, 3, 3]
	Temporal Convolution_1	Conv2d: (1,3)	[20, 90, 1, 65]
	Graph Convolution		[20, 90, 30]
	Temporal Convolution_2	Conv2d: (1,3)	[20, 90, 1, 30]
Classification Module	Uncertainty Fusion		[20, 90, 30]
	Flatten		[20, 2700]
	Linear Layer_1		[20, 1024]
	BatchNorm1d		[20, 1024]
	Dropout		[20, 1024]
	Linear Layer_2		[20, 256]
	BatchNorm1d		[20, 256]
	Dropout		[20, 256]
	Linear Layer_3		[20, 2]

* "B" means "batch size"; "T" means "the number of windows"; "N" means "the number of nodes"; "D" means "dimensions of feature"

classification accuracy across 25 subjects. While many existing methods are tailored specifically for fMRI data, excelling in leveraging its high spatial resolution and relatively stable connectivity patterns, they often lack adaptability to other neural data modalities such as EEG. In contrast, our method, though originally designed for fMRI, demonstrated superior performance on the motor imagery (MI) EEG dataset compared to methods specifically developed for fMRI-based analysis.

Unlike fMRI-specific approaches that focus on static or coarse-grained spatial relationships, our model integrates spatio-temporal convolution and temporal attention mechanisms. This design effectively captures the rapidly changing temporal dynamics inherent in EEG data, which are critical for motor imagery (MI) tasks. Moreover, our method incorporates a confidence-weighted feature fusion mechanism that addresses signal quality variations across time windows. This mechanism mitigates the impact of noisy or less informative EEG segments, thereby enhancing the model's robustness.

B. Framework Details

A.Table. V provides a detailed description of the proposed framework, including the architecture components, parameter settings, and additional configurations used in the experiments. This supplementary information aims to offer a comprehensive understanding of the framework implementation.

C. Evaluation Metrics Analysis

To comprehensively evaluate the model's performance across different categories, we calculate both micro-level and macro-level evaluation metrics (including ACC and F1). The experimental results in A.Table. VI show that the micro-level and macro-level values (ACC and F1) are very similar across both datasets. As detailed in Section IV, the sample distribution of the datasets used in the experiment is relatively balanced, which helps minimize the difference between the micro-level and macro-level calculations. This also indicates

TABLE VI
MICRO-LEVEL AND MACRO-LEVEL ANALYSIS (%)

ADNI	NC vs. SMC&EMCI		NC vs. SMC		NC vs. EMCI		SMC vs. EMCI	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1
FBNetGen	84.33 / 83.00	88.74 / 87.96	78.76 / 80.78	74.98 / 72.89	84.89 / 84.20	76.96 / 76.37	66.92 / 68.76	53.12 / 56.13
GraphSAGE	85.66 / 85.64	88.49 / 88.08	86.71 / 88.14	83.93 / 83.75	89.63 / 89.12	**87.81** / 86.67	71.53 / 72.30	60.96 / 61.58
SAGPool	85.21 / 85.16	88.65 / 85.24	85.33 / 86.95	81.01 / 82.72	87.85 / 87.39	76.34 / 78.24	70.00 / 71.07	63.49 / 65.63
FE-STGNN	82.46 / 83.14	87.53 / 87.04	81.25 / 80.88	84.32 / 84.10	82.02 / 82.16	80.00 / 80.67	74.56 / 74.57	72.00 / 71.42
ST-fMRI	87.15 / 88.03	89.36 / 87.84	86.27 / 86.53	85.09 / 85.71	88.04 / 87.91	83.45 / 83.76	76.02 / 74.92	73.28 / 72.38
ST-GCN	85.83 / 87.73	87.42 / 86.69	88.14 / 85.52	85.98 / 84.71	84.06 / 82.97	82.23 / 81.52	72.46 / 73.84	69.38 / 68.91
Ours	**88.21 / 88.69**	**89.86 / 88.84**	**89.64 / 90.42**	**86.87 / 88.84**	**89.85 / 89.65**	87.68 / **87.41**	**78.23 / 78.46**	**75.00 / 75.24**
PD	NC vs. TDPD&PIGDPD		NC vs. TDPD		NC vs. PIGDPD		TDPD vs. PIGDPD	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1
FBNetGen	80.84 / 78.46	80.17 / 79.48	74.39 / 74.54	72.50 / 73.36	78.67 / 78.66	65.99 / 65.32	75.90 / 76.54	60.53 / 61.68
GraphSAGE	77.35 / 76.91	80.24 / 78.39	77.87 / 77.45	75.69 / 75.95	77.77 / 76.33	74.07 / 72.99	74.09 / 74.18	60.18 / 61.72
SAGPool	76.54 / 75.14	82.11 / 81.27	73.71 / 73.78	71.24 / 70.47	78.55 / 80.55	71.53 / 72.52	76.90 / 76.00	59.63 / 60.72
FE-STGNN	73.86 / 73.57	79.28 / 80.32	72.80 / 73.77	70.25 / 70.68	75.73 / 75.70	65.07 / 65.86	68.77 / 68.50	64.25 / 64.21
ST-fMRI	76.12 / 77.76	78.06 / 78.83	72.79 / 72.92	67.33 / 70.15	73.84 / 73.23	67.06 / 70.92	75.32 / 75.61	64.33 / 62.34
ST-GCN	80.11 / 80.68	82.65 / 80.67	76.89 / 78.78	78.56 / 78.37	80.88 / 82.55	74.12 / 75.67	73.36 / 69.63	50.25 / 47.08
Ours	**81.20 / 80.88**	**83.55 / 83.77**	**82.71 / 81.53**	**83.76 / 83.78**	**84.77 / 84.66**	**76.72 / 76.93**	**77.00 / 76.73**	**65.61 / 67.01**

* Micro-level evaluation (%) / Macro-level evaluation (%)

that our model performs consistently across different categories.

REFERENCES

[1] J. M. Shine, "The thalamus integrates the macrosystems of the brain to facilitate complex, adaptive brain network dynamics," *Prog. Neurobiol.*, vol. 199, Apr. 2021, Art. no. 101951.

[2] Y. Yang, H. Cui, and C. Yang, "Pre-train graph neural networks for brain network analysis," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2022, pp. 4993–4994.

[3] R. J. Deligani et al., "Electrical and hemodynamic neural functions in people with ALS: An EEG-fNIRS resting-state study," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 3129–3139, Dec. 2020.

[4] J. M. Ranson et al., "Harnessing the potential of machine learning and artificial intelligence for dementia research," *Brain Informat.*, vol. 10, no. 1, p. 6, Dec. 2023.

[5] Y. Ma et al., "Multi-scale dynamic graph learning for brain disorder detection with functional MRI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3501–3512, 2023.

[6] X. Sun, L. Liu, and C. Ye, "Advances in functional magnetic resonance imaging," *Chin. J. Neurosci.*, vol. 17, no. 3, pp. 270–272, 2001.

[7] Y. Yang, C. Ye, X. Guo, T. Wu, Y. Xiang, and T. Ma, "Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning," *IEEE Trans. Med. Imag.*, vol. 43, no. 1, pp. 108–121, Jan. 2024.

[8] H. Zhang et al., "Classification of brain disorders in rs-fMRI via local-to-global graph neural networks," *IEEE Trans. Med. Imag.*, vol. 42, no. 2, pp. 444–455, Feb. 2023.

[9] H. Ju and D. S. Bassett, "Dynamic representations in networked neural systems," *Nature Neurosci.*, vol. 23, no. 8, pp. 908–917, Aug. 2020.

[10] J. Liu, X. H. Liao, M. R. Xia, and Y. He, "Chronnectome fingerprinting: Identifying individuals and predicting higher cognitive functions using dynamic brain connectivity patterns," *Hum. Brain Mapp.*, vol. 39, no. 2, pp. 902–915, 2018.

[11] B. Ge, H. Wang, P. Wang, Y. Tian, X. Zhang, and T. Liu, "Discovering and characterizing dynamic functional brain networks in task FMRI," *Brain Imag. Behav.*, vol. 14, no. 5, pp. 1660–1673, Oct. 2020.

[12] Y. Zhang and C. Zhu, "Assessing brain networks by resting-state dynamic functional connectivity: An fNIRS-EEG study," *Frontiers Neurosci.*, vol. 13, p. 1430, Jan. 2020.

[13] B. Jie, M. Liu, and D. Shen, "Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease," *Med. Image Anal.*, vol. 47, pp. 81–94, Jul. 2018.

[14] F. Noman, S.-Y. Yap, R. C.-W. Phan, H. Ombao, and C.-M. Ting, "Graph autoencoder-based embedded learning in dynamic brain networks for autism spectrum disorder identification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 2891–2895.

[15] A. Zou, J. Ji, M. Lei, J. Liu, and Y. Song, "Exploring brain effective connectivity networks through spatiotemporal graph convolutional models," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 6, pp. 7871–7883, Jun. 2024.

[16] H. Huang, Q. Liu, Y. Jiang, Q. Yang, X. Zhu, and Y. Li, "Deep spatio-temporal attention-based recurrent network from dynamic adaptive functional connectivity for MCI identification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2600–2612, 2022.

[17] X. Zhao et al., "Deep reinforcement learning guided graph neural networks for brain network analysis," *Neural Netw.*, vol. 154, pp. 56–67, Oct. 2022.

[18] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*.

[19] Q. Zhu, C. Zheng, Z. Zhang, W. Shao, and D. Zhang, "Dynamic confidence-aware multi-modal emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 15, no. 3, pp. 1358–1370, Jul. 2024.

[20] M. Bahrami, P. J. Laurienti, H. M. Shappell, D. Dagenbach, and S. L. Simpson, "A mixed-modeling framework for whole-brain dynamic network analysis," *Netw. Neurosci.*, vol. 6, no. 2, pp. 591–613, Jun. 2022.

[21] F. Tong and M. S. Pratte, "Decoding patterns of human brain activity," *Annu. Rev. Psychol.*, vol. 63, pp. 483–509, Jan. 2012.

[22] W. J. Jagust and S. Landau, "Temporal dynamics of β-amyloid accumulation in aging and Alzheimer disease," *Neurology*, vol. 96, no. 9, pp. 1347–1357, Jan. 2021.

[23] R. Xu, Q. Zhu, S. Li, Z. Hou, W. Shao, and D. Zhang, "MSTGC: Multi-channel spatio-temporal graph convolution network for multi-modal brain networks fusion," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2359–2369, 2023.

[24] E. Damaraju et al., "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *NeuroImage, Clin.*, vol. 5, pp. 298–308, Jan. 2014.

[25] S. Dahan, L. Z. Williams, D. Rueckert, and E. C. Robinson, "Improving phenotype prediction using long-range spatio-temporal dynamics of functional connectivity," in *Proc. 4th Int. Workshop, MLCN, Held Conjunct MICCAI*. Strasbourg, France: Springer, Sep. 2021, pp. 145–154.

[26] B.-H. Kim, J. C. Ye, and J.-J. Kim, "Learning dynamic graph representation of brain connectome with spatio-temporal attention," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 4314–4327.

[27] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: A comprehensive review," *Comput. Social Netw.*, vol. 6, no. 1, pp. 1–23, Dec. 2019.

[28] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2021.

[29] X. Li, L. Sun, M. Ling, and Y. Peng, "A survey of graph neural network based recommendation in social networks," *Neurocomputing*, vol. 549, Sep. 2023, Art. no. 126441.

[30] H. Cai, H. Zhang, D. Zhao, J. Wu, and L. Wang, "FP-GNN: A versatile deep learning architecture for enhanced molecular property prediction," *Briefings Bioinf.*, vol. 23, no. 6, p. 408, Nov. 2022.

[31] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[32] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1025–1035.

[33] S. I. Ktena et al., "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, Apr. 2018.

[34] J.-H. Choi and J.-S. Lee, "Confidence-based deep multimodal fusion for activity recognition," in *Proc. ACM Int. Joint Conf. Int. Symp. Pervasive Ubiquitous Comput. Wearable Comput.*, Oct. 2018, pp. 1548–1556.

[35] G. Deng, K. Zou, M. Wang, X. Yuan, S. Ying, and H. Fu, "A multi-view impartial decision network for frontotemporal dementia diagnosis," 2023, *arXiv:2307.04981*.

[36] Z. Song et al., "EEG-MACS: Manifold attention and confidence stratification for EEG-based cross-center brain disease diagnosis under unreliable annotations," in *Proc. 32nd ACM Int. Conf. Multimedia*, Oct. 2024, pp. 340–349.

[37] S.-Y. Shih, F.-K. Sun, and H.-Y. Lee, "Temporal pattern attention for multivariate time series forecasting," *Mach. Learn.*, vol. 108, nos. 8–9, pp. 1421–1441, Sep. 2019.

[38] J. Hauke and T. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data," *Quaestiones Geographicae*, vol. 30, no. 2, pp. 87–93, Jun. 2011.

[39] L. Qiao, H. Zhang, M. Kim, S. Teng, L. Zhang, and D. Shen, "Estimating functional brain networks by incorporating a modularity prior," *NeuroImage*, vol. 141, pp. 399–407, Nov. 2016.

[40] X. Kan, H. Cui, J. Lukemire, Y. Guo, and C. Yang, "FBNETGEN: Task-aware GNN-based fMRI analysis via functional brain network generation," in *Proc. Int. Conf. Med. Imag. Deep Learn.*, Jan. 2022, pp. 618–637.

[41] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3734–3743.

[42] H. Zhang, X. Chen, Y. Zhang, and D. Shen, "Test-retest reliability of 'high-order' functional connectivity in young healthy adults," *Frontiers Neurosci.*, vol. 11, p. 439, Aug. 2017.

[43] Y. Zhang, H. Zhang, X. Chen, S.-W. Lee, and D. Shen, "Hybrid high-order functional connectivity networks using resting-state functional MRI for mild cognitive impairment diagnosis," *Sci. Rep.*, vol. 7, no. 1, p. 6530, Jul. 2017.

[44] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[45] S. S. Menon and K. Krishnamurthy, "A comparison of static and dynamic functional connectivities for identifying subjects and biological sex using intrinsic individual brain connectivity," *Sci. Rep.*, vol. 9, p. 5729, Apr. 2019.

[46] R. Hu, L. Peng, J. Gan, X. Shi, and X. Zhu, "Complementary graph representation learning for functional neuroimaging identification," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 3385–3393.

[47] D. Chen and L. Zhang, "FE-STGNN: Spatio-temporal graph neural network with functional and effective connectivity fusion for MCI diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Lima, Peru: Springer, 2023, pp. 67–76.

[48] X. Chen et al., "High-order resting-state functional connectivity network for MCI classification," *Hum. Brain Map.*, vol. 37, no. 9, pp. 3282–3296, 2016.

[49] S. Gadgil et al., "Spatio-temporal graph convolution for resting-state fMRI analysis," in *Proc. 23rd Int. Conf.* Vancouver, BC, Canada: Springer, Oct. 2020, pp. 528–538.

[50] Z. Zhou et al., "A toolbox for brain network construction and classification (BrainNetClass)," *Hum. Brain Mapping*, vol. 41, no. 10, pp. 2808–2826, Jul. 2020.

[51] M. P. Richardson, B. A. Strange, J. S. Duncan, and R. J. Dolan, "Preserved verbal memory function in left medial temporal pathology involves reorganisation of function to right medial temporal lobe," *NeuroImage*, vol. 20, pp. 112–119, Nov. 2003.

[52] P. Liang, Z. Wang, Y. Yang, and K. Li, "Three subsystems of the inferior parietal cortex are differently affected in mild cognitive impairment," *J. Alzheimer's Disease*, vol. 30, no. 3, pp. 475–487, Jun. 2012.

[53] S. Jacobson, E. M. Marcus, S. Pugsley, S. Jacobson, E. M. Marcus, and S. Pugsley, "The limbic system, temporal lobe, and prefrontal cortex," in *Neuroanatomy for the Neuroscientist*, Jan. 2018, pp. 477–529. [Online]. Available: https://doi.org/10.1007/978-3-319-60187-8_16

[54] S. A. R. B. Rombouts, F. Barkhof, R. Goekoop, C. J. Stam, and P. Scheltens, "Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: An fMRI study," *Hum. Brain Mapping*, vol. 26, no. 4, pp. 231–239, Dec. 2005.

[55] Y. Gu et al., "Abnormal dynamic functional connectivity in Alzheimer's disease," *CNS Neurosci. Therapeutics*, vol. 26, no. 9, pp. 962–971, 2020.

[56] J. Ma, B. Yang, W. Qiu, Y. Li, S. Gao, and X. Xia, "A large EEG dataset for studying cross-session variability in motor imagery brain-computer interface," *Sci. Data*, vol. 9, no. 1, p. 531, Sep. 2022.