NeuroImage 307 (2025) 121013

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Multi-channel spatio-temporal graph attention contrastive network for brain disease diagnosis

Chaojun Li [a,b], Kai Ma [c,*], Shengrong Li [a,b], Xiangshui Meng [d], Ran Wang [e], Daoqiang Zhang [a,b], Qi Zhu [a,b,*]

[a] College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China
[b] Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China
[c] School of Computer, Jiangsu University of Science and Technology, Zhenjiang, 212100, China
[d] Department of Radiology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, Qingdao, 266035, China
[e] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China

ARTICLE INFO

Dataset link: http://adni.loni.usc.edu/, https://github.com/xbrainnet/MSTGAC.git

Keywords:
Brain disease diagnosis
Dynamic brain networks
Spatio-temporal features
Graph contrastive learning

ABSTRACT

Dynamic brain networks (DBNs) can capture the intricate connections and temporal evolution among brain regions, becoming increasingly crucial in the diagnosis of neurological disorders. However, most existing researches tend to focus on isolated brain network sequence segmented by sliding windows, and they are difficult to effectively uncover the higher-order spatio-temporal topological pattern in DBNs. Meantime, it remains a challenge to utilize the structure connectivity prior in the DBNs analysis. To address these problems, we propose a multi-channel spatio-temporal graph attention contrastive network for DBNs analysis. Specifically, we first construct dynamic brain functional networks from fMRI data with sliding windows, and embed the structural connectivity derived from diffusion tensor imaging (DTI) to the dynamic functional connectivity graph representation to construct multi-modal brain network. Second, we develop a multi-channel spatial attention contrastive network to extract topological features from the brain network within each time window. This network incorporates an intra-window graph contrastive constraint to enhance the discriminative ability of the extracted features. Moreover, temporal dependencies across windows are captured by integrating feature embeddings through a self-attention mechanism, and the inter-window recurrent contrastive constraint is devised to extract higher-order spatio-temporal topological features. Finally, a multi-layer perceptron (MLP) is used to classify the brain networks. Experiments on epilepsy and ADNI datasets show that our method outperforms several state-of-the-art approaches in diagnosing performance, and it provides discriminative graph features for related brain diseases.

1. Introduction

The brain is a complex system composed of neurons and the intricate network of connections interlacing them. The highly interconnected connection structure in brain is often abstracted as brain networks (Bullmore and Sporns, 2009). In recent years, brain networks analysis has become a popular tool in neuroscience research and diagnosis of neurological disorders, as it is capable of revealing the complex topological relationships and temporal dynamics between brain regions (Van Den Heuvel and Sporns, 2011). Previous studies have demonstrated that changes in the activation state and connection pattern of the brain are implicated in a wide variety of neurodegenerative diseases, such as schizophrenia, Alzheimer's disease, and epilepsy. Márquez and Yassa (2019), Drenthen et al. (2020). Therefore, brain networks provides an effective tool for identifying brain diseases and biomarkers.

Functional magnetic resonance imaging (fMRI) is considered the essential tool in investigating brain functional networks (Yu et al., 2012). It reflects a pattern of activation across different brain regions in certain functional cognition behavior by detecting the fluctuation of blood flow and oxygenation. The connectivity among different brain regions can be quantified by calculating the statistical dependencies or neural activity time series correlations from different parts of the brain, which is often based on a metric such as Pearson's correlation or coherence. For example, Van Den Heuvel and Sporns (2011) used fMRI data from a large dataset to apply Pearson correlation for constructing whole brain functional connectomes. Similarly, Smith et al. (2013)

* Corresponding authors.
E-mail addresses: lichaojun@nuaa.edu.cn (C. Li), kaim@nuaa.edu.cn (K. Ma), lisrong@nuaa.edu.cn (S. Li), qdqlfsk2021@163.com (X. Meng), wangran@nuaa.edu.cn (R. Wang), dqzhang@nuaa.edu.cn (D. Zhang), zhuqinuaa@163.com (Q. Zhu).

https://doi.org/10.1016/j.neuroimage.2025.121013
Received 14 October 2024; Received in revised form 4 January 2025; Accepted 8 January 2025
Available online 14 January 2025

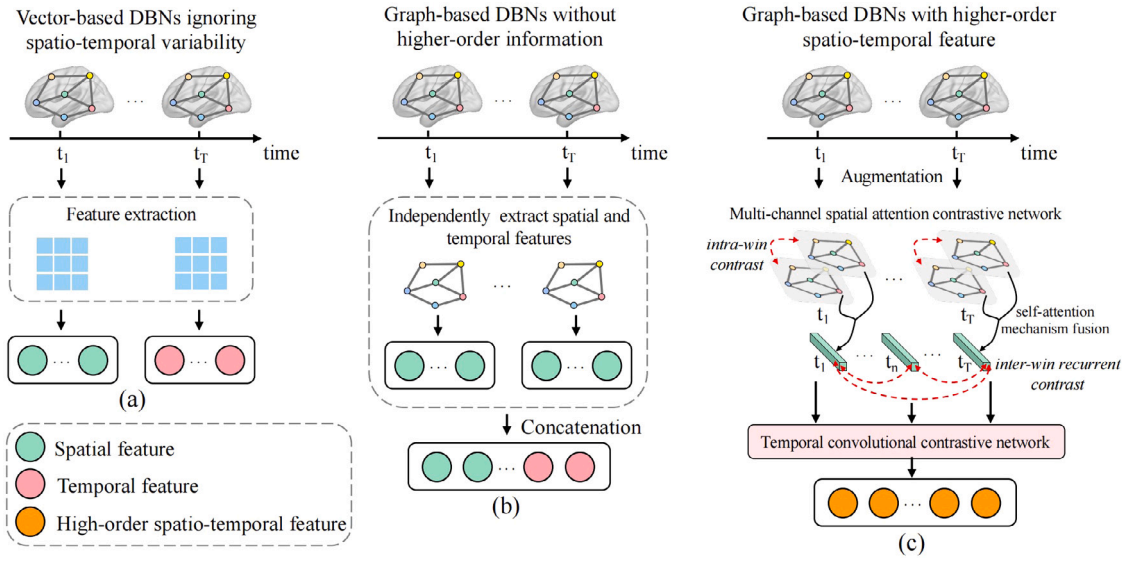


Fig. 1. A comparison of the proposed method with existing brain network analysis techniques: (a) vector-based method, (b) conventional graph-based method, and (c) the proposed method.

applied independent component analysis to static fMRI data from the Human Connectome Project to identify resting-state networks, revealing consistent patterns of functional connectivity across individuals. However, these brain networks methods often regard the functional brain networks as static networks, focusing on the mean functional connectivity during the scanning session and frequently overlooking the dynamic temporal fluctuations in connectivity (Zalesky et al., 2014).

In contrast to static brain networks, DBNs elucidate the temporal dynamics of functional connections between various brain regions, thereby capturing the evolution of brain functional connectivity over time. For examples, Chung et al. (2024) employed dynamic topological data analysis methods to reveal the mechanism by which brain networks facilitate brain function and adaptive behavior. Zhang et al. (2020) explored the dynamic reorganization of brain network functional topology as it transitions from a resting state to a task-engaged state. Liu (2024) conducted a topological analysis to reveal the overlapping structural characteristics of brain functional networks under natural observation conditions. Gai et al. (2023) combined temporal and spatial variability features from fMRI and applied a random forest model to classify major depressive disorder. Kottaram et al. (2018) identified the individual differences in schizophrenia diagnosis by analyzing the spatio-temporal dynamics of brain networks. It is worth noting that the previous DBNs analysis methods are limited to extracting features that are isolated within discrete temporal segments, and struggle to comprehensively exploit the intricate and higher-order spatio-temporal topological characteristics across extended temporal intervals.

Furthermore, most of the DBNs analysis methods are predominantly concerned with the fluctuating graph features of functional connectivity. However, they frequently fail to account for the substantial influence that structural connectivity, typically assessed via DTI, exerts on the DBNs throughout the analytical process. Neuroscientific research indicates that the heterogeneity of structural connections enables the brain to process information diversely, fostering specialization and integration among different brain regions, which may consequently have an impact on the connectivity patterns within DBNs (Batista-García-Ramó and Fernández-Verdecia, 2018). Honey et al. (2010) explored the relationship between functional and structural connectivity within the human brain, revealing that structural networks, particularly white matter pathways, are strong predictors of functional connectivity patterns. Likewise, Hagmann et al. (2008) emphasized that brain structure governs the flow of neural activity, underscoring the importance of integrating structural data into analysis of DBNs. Incorporating structural connectivity is crucial for providing a comprehensive understanding of how dynamic functional patterns arise and evolve. Structural information in the DBNs propagates concurrently in time and space, and maintains a close bond of global interaction across different temporal intervals. Without considering these anatomical connections, it is difficult to effectively grasp how disruptions in brain networks contribute to the progression of neurological diseases and associated cognitive impairments.

Fig. 1 provides a comparative analysis of our proposed method and existing approaches in DBNs analysis. From Fig. 1(a), we can find that the vector-based DBNs analysis methods overlook the spatio-temporal topological variability of brain networks. It isolates the extraction of temporal and spatial features, disrupting the topological structure and making it challenging to effectively capture the spatio-temporal topological information. As shown in Fig. 1(b), the graph-based DBNs analysis methods can preserve the topological structure and extract time and space features separately. But they fail to capture higher-order spatio-temporal information in DBNs. In Fig. 1(c), our proposed method can not only offer a unified framework for extracting spatio-temporal topological information from DBNs but also thoroughly explore the discriminative higher-order information within and across time windows.

To address the above issues, we propose a multi-channel spatio-temporal graph attention contrastive network (MSTGAC) to extract spatio-temporal topological information from DBNs for brain disease diagnosis. First, we segment fMRI time series with sliding windows, and then construct multi-modal brain networks by fusing functional connectivity matrices from each time window and structural connectivity matrices. Second, we develop a multi-channel spatial attention contrastive network that is capable of extracting topological features from brain networks for each specific time segment. This network includes a graph contrastive constraint within the time window to improve the discriminability of the features. Additionally, we propose the inter-window recurrent contrastive constraint to capture higher-order spatio-temporal features across time windows. Ultimately, the features are then classified through an MLP. The main insights from this study are outlined below.

1. In this paper, we propose a deep graph feature extraction model for DBNs analysis, which facilitates the propagation of higher-order spatio-temporal topological information in DBNs.

2. We propose a multi-modal brain networks analysis method that seamlessly integrates structural and functional connectivity into a unified graph representation, effectively embedding prior knowledge of structural connectivity into the feature extraction process of DBNs.
3. We develop a multi-channel spatial attention network equipped with both intra-window and inter-window contrastive constraints. These constraints enhance the discriminative power of spatio-temporal topological features within and across time windows.
4. Experimental results on epilepsy and ADNI datasets demonstrate that our proposed method outperforms several state-of-the-art brain networks analysis methods, which highlights its potential for identifying biomarkers related to neurological disorders.

2. Methodology

In this study, we introduce the MSTGAC to effectively classify DBNs, and apply it to brain disorder diagnosis. The structure of the proposed model is depicted in Fig. 2, and a detailed explanation follows:

2.1. Augmentation of dynamic brain networks

In this work, we first generate DBNs for each participant by sliding windows method. Specifically, for each participant, the BOLD time series signals from the rs-fMRI data are denoted as $X = (x_1, x_2, \dots, x_N)^T \in \mathbb{R}^{N\times M}$, where N is the number of segmented regions of interests (ROIs), and M denotes the number of continuous time series points. We divide all rs-fMRI time series into T sub-sequence segments by non-overlapping sliding windows with length L. For the N brain regions, the t^{th} sub-sequence segment is represented by a matrix $x(t) = (x_1(t), x_2(t), \dots, x_N(t))^T \in \mathbb{R}^{N\times L}$, where $x_i(t) \in \mathbb{R}^L$, $i = (1, 2, \dots, N)$, $t = (1, 2, \dots, T)$, and $x_i(t)$ represents the t^{th} sub-sequence segment extracted from the rs-fMRI time series of the i^{th} ROI. We generate the functional connectivity network by computing the Pearson correlation coefficients between the i^{th} and j^{th} brain regions at the t^{th} time interval, which provides a quantitative measure of the functional connection between the two regions.

$$c_{ij}(t) = \frac{cov(x_i(t), x_j(t))}{\sigma_{x_i(t)}\sigma_{x_j(t)}} \tag{1}$$

where $cov(x_i(t), x_j(t))$ represents the covariance of $x_i(t)$ and $x_j(t)$, and σ denotes the standard deviation. We derive the dynamic functional connectivity network $C = (c(1), c(2), \dots, c(T))^T \in \mathbb{R}^{T\times N\times N}$, where $c(t)$ denotes the functional connectivity network corresponding to the t^{th} time segment, and each element of $c(t)$ is given by $c_{ij}(t)$. Additionally, the structural connectivity network is represented as $D = (d_1, d_2, \dots, d_N)^T \in \mathbb{R}^{N\times N}$, where the value d_i indicates the intensity of connections between brain regions. To capture cross-modal interaction information between fMRI and DTI data, we define the DBNs for each subject as $G = (g(1), g(2), \dots, g(T))^T$, with $g(t) = (c(t), D)$, where $c(t) \in \mathbb{R}^{N\times N}$ serves as the feature matrix of nodes in a brain network (e.g., $g(t)$), and $D \in \mathbb{R}^{N\times N}$ denotes as the adjacency matrix of a brain network. Importantly, the graph structure $g(t)$ is subject-dependent, meaning that both the feature matrix $c(t)$ and the adjacency matrix D are individually constructed for each subject based on their unique fMRI and DTI data. This subject-specific design ensures that the generated graph fully captures the unique characteristics of each individual's brain network, allowing our method to effectively model personalized cross-modal interactions.

Physiological and pathological changes in the brain may lead to the temporal and spatial variations in the connectivity patterns of brain networks. To improve the proposed model's stability in diagnosing brain diseases, we adopt the contrastive learning to generate two views for data augmentation at the graph domain level of the brain network $g(t)$ at t^{th} time window. We design a unified method to corrupt both node features and topological connection features by randomly deleting valid values from the feature matrix $c(t)$ and the adjacency matrix D, enabling contrastive learning at both the feature and topology levels. For the augmentation of the feature matrix, we first generate a random masking matrix $R \in \{0, 1\}^{N\times N}$, where each entry is sampled from a Bernoulli distribution $R_{ij} \sim \mathcal{B}(1 - p_{f1})$, where p_{f1} is the probability of valid values being removed. The resulting feature can be computed as:

$$\widetilde{c^1(t)} = c(t)\circ R \tag{2}$$

where $\circ$ is Hadamard product. Similarly, for the augmentation of the adjacency matrix, we generate a random masking matrix $Q \in \{0, 1\}^{N\times N}$, where each element Q_{ij} is independently sampled from a Bernoulli distribution $Q_{ij} \sim \mathcal{B}(1 - p_{a1})$. Here, p_{a1} denotes the probability that an edge is removed. The resulting adjacency matrix is then expressed as:

$$\widetilde{D}^1 = D\circ Q \tag{3}$$

We generate two contrastive views $g^1(t) = (\widetilde{c^1(t)}, \widetilde{D}^1)$ and $g^2(t) = (\widetilde{c^2(t)}, \widetilde{D}^2)$ using a uniform augmentation approach, where their corruption levels are controlled by two different sets of hyperparameters: p_{f1}, p_{a1} and p_{f2}, p_{a2}.

2.2. Multi-channel spatio-temporal graph attention contrastive network

We design the MSTGAC to capture higher-order features from brain networks under spatio-temporal variations. Specifically, we first employ a multi-channel spatial attention contrastive network to extract complex topological information from DBNs within each time window, and apply an intra-window contrastive optimization objective to capture discriminative features. Subsequently, a temporal convolutional contrastive network is used to extract dynamic temporal information, using an inter-window contrastive optimization objective to explore higher-order discriminative information across time windows.

2.2.1. Multi-channel spatial attention contrastive network

Based on the augmented DBNs, each time window contains two brain networks at different levels of corruption. To capture the complexity of the topological structure of the brain networks, we use a shared-weight multi-channel graph attention network (GAT) (Velickovic et al., 2017) to extract spatial information from brain networks (i.e., $g(t)$) with different corruption levels within the same time window, as shown in Fig. 2. Specifically, given a view $g^1(t) = (\widetilde{c^1(t)}, \widetilde{D}^1)$, the attention coefficient between brain regions i and j is defined as:

$$h_{ij} = LeakyReLU(a^T[W\widetilde{c_i^1(t)} \parallel W\widetilde{c_j^1(t)}]) \tag{4}$$

where W denotes a learnable weight matrix, $a \in \mathbb{R}^{2N}$ represents a learnable attention vector, $LeakyReLU$ is the LeakyReLU activation function, and the symbol $\parallel$ denotes the concatenation operation. Besides, $\widetilde{c_i^1(t)}$ and $\widetilde{c_j^1(t)}$ are the features of i^{th} and j^{th} brain regions of dynamic functional connectivity network in the t^{th} time window. We use a mask mechanism to ensure the attention coefficient incorporates the graph structure, which is defined as:

$$\widetilde{H} = H\circ D \tag{5}$$

where $H \in \mathbb{R}^{N\times N}$ is the attention coefficient matrix, $D \in \mathbb{R}^{N\times N}$ denotes the adjacency matrix, $\widetilde{H} \in \mathbb{R}^{N\times N}$ represents the attention coefficient matrix incorporating the graph structure, and the value of matrix $\widetilde{H}$ is $\widetilde{h_{ij}}$. We normalize $\widetilde{h_{ij}}$ across neighboring brain region $j \in N_i$ of the i^{th} brain region using the SoftMax function, where N_i is the number of adjacent brain regions of i^{th} brain region.

$$\alpha_{ij} = \frac{\exp(\widetilde{h_{ij}})}{\sum_{k\in N_i}\exp(\widetilde{h_{ik}})} \tag{6}$$

Then, the coefficients are employed to compute a weighted sum of the corresponding feature values in order to produce the output feature

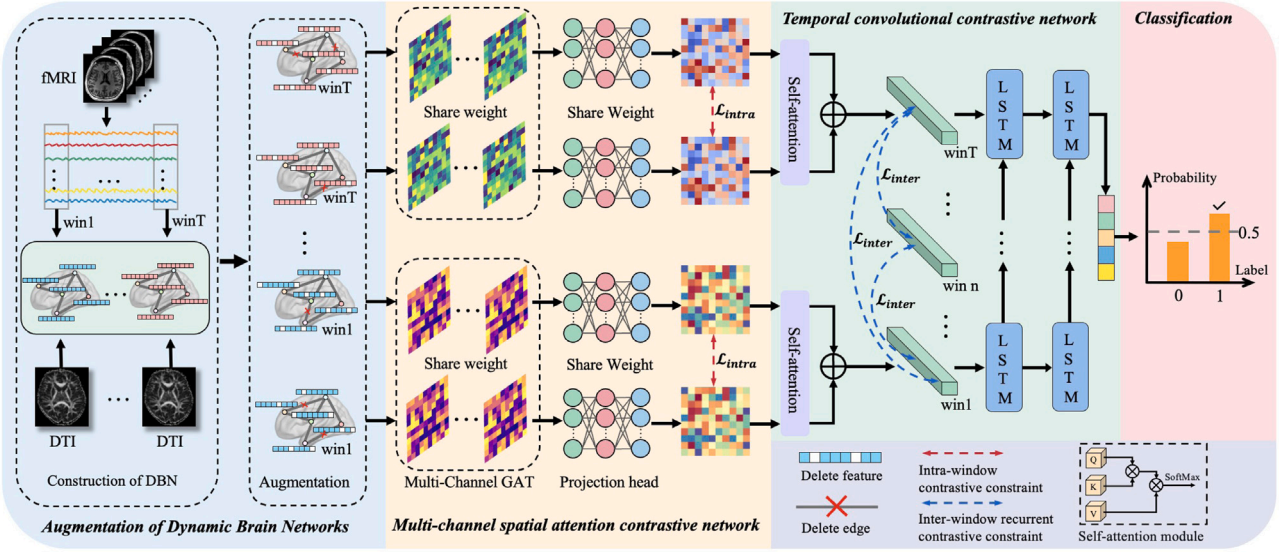


Fig. 2. The architecture of the proposed multi-channel spatio-temporal graph attention contrastive network for diagnosing brain diseases.

for each node:

$$\widetilde{c_i^1(t)}' = \sum_{j \in N_i} \alpha_{ij} W \widetilde{c_i^1(t)} \tag{7}$$

Through the multi-channel GAT, the features of the two contrastive views at t^{th} time window are transformed into $g'^1(t) \in \mathbb{R}^{N \times N}$ and $g'^2(t) \in \mathbb{R}^{N \times N}$. Subsequently, we use a projection head as a nonlinear transformation to map the feature representations into the latent feature space (Chen et al., 2020). In this paper, we apply a two-layer perceptron on $g'^1(t)$ and $g'^2(t)$ to obtain the features $z^1(t) \in \mathbb{R}^{N \times N}$ and $z^2(t) \in \mathbb{R}^{N \times N}$ after passing through the projection head. In the latent feature space, we design an intra-window contrastive optimization objective to capture the features that are discriminative for disease diagnosis. Specifically, we denote the embeddings of the i^{th} brain region in $z^1(t)$ as p_i, which is considered as the anchor sample. Oppositely, the embeddings of the i^{th} brain region in $z^2(t)$ are considered as the positive sample, represented as q_i. Naturally, the embeddings in $z^2(t)$ except q_i are regarded as negative samples. We define the contrastive optimization objective for each positive pair (p_i, q_i) as:

$$\eta(p_i, q_i) = -log \frac{e^{cos(p_i, q_i)/\tau}}{e^{cos(p_i, q_i)/\tau} + \sum_{k=1, k \neq i}^{N} e^{cos(p_i, q_k)/\tau}} \tag{8}$$

where $\cos(\cdot)$ denotes the cosine similarity, and τ represents the temperature parameter.

2.2.2. Temporal convolutional contrastive network

The feature embeddings $z^1(t)$ and $z^2(t)$ enable the extraction of internal spatial specificity from the contrastive views using self-attention mechanism. We apply different transformations to the embeddings using linear mapping:

$$Q^1(t) = z^1(t)W_Q^1(t), \quad K^1(t) = z^1(t)W_K^1(t), \quad V^1(t) = z^1(t)W_V^1(t) \tag{9}$$

$$Q^2(t) = z^2(t)W_Q^2(t), \quad K^2(t) = z^2(t)W_K^2(t), \quad V^2(t) = z^2(t)W_V^2(t) \tag{10}$$

where $W_Q(t), W_K(t), W_V(t) \in \mathbb{R}^{N \times N}$ represent the parameter matrices for generating queries, keys, and values, respectively, which are updated through backpropagation. The attention coefficients in the self-attention mechanism are calculated by performing a scaled dot-product between Q and K, followed by applying the SoftMax function. The self-attention feature matrix is obtained by multiplying the attention weights with V. In other words, it is calculated by fusing the self-attention feature matrix from the contrastive views:

$$s^1(t) = softmax\left(\frac{Q^1(t)K^1(t)}{\sqrt{\gamma^1}}\right)V^1(t),$$

$$s^2(t) = softmax\left(\frac{Q^2(t)K^2(t)}{\sqrt{\gamma^2}}\right)V^2(t) \tag{11}$$

$$s(t) = s^1(t) + s^2(t) \tag{12}$$

where $s^1(t), s^2(t) \in \mathbb{R}^{N \times N}$ denote the self-attention feature matrices, $s(t) \in \mathbb{R}^{N \times N}$ represents the output of the backbone network, and γ^1, γ^2 are the normalization parameter, which are equal to the dimension of K. After passing the contrastive views by the self-attention mechanism, we obtain feature embedding $S = (s(1), s(2), \ldots, s(T))^T$, which contains rich temporal information.

There are many important potential connections in the transient brain networks in time dimension, which is crucial for brain disease diagnosis to capture the dynamic evolution of brain networks. We design an inter-window contrastive optimization objective to capture the potential associations from different time windows. Specifically, the embedding of the i^{th} brain region at the m^{th} time window is defined as the anchor sample $u_i(m)$, and the embedding of the i^{th} brain region at the n^{th} time window is defined as the positive sample $v_i(n)$. Likewise, the definition of negative samples is the same as the intra-window contrastive optimization objective. The contrastive optimization objective for each positive pair $(u_i(m), v_i(n))$ can be calculated as:

$$\eta(u_i(m), v_i(n)) = -log \frac{e^{cos(u_i(m), v_i(n))/\tau}}{e^{cos(u_i(m), v_i(n))/\tau} + \sum_{k=1, k \neq i}^{N} e^{cos(u_i(m), v_k(n))/\tau}} \tag{13}$$

To avoid the gradient vanishing and gradient explosion in multi-channel spatial attention contrastive network, we use a two-layer stacked LSTM (Hochreiter, 1997) unit to capture the dynamic temporal information transmitted across time windows in $S = (s(1), s(2), \ldots, s(T))^T$, applying batch normalization and *tanh* activation after each LSTM layer. Finally, an MLP is utilized to capture the relationship between the dynamic feature embeddings and the predicted outcomes.

2.3. Time windows-based dual contrastive loss function

2.3.1. Intra-window contrastive constraint

In the multi-channel spatial attention contrastive network, we propose an optimization objective $\eta(p_i, q_i)$ to explore the discriminative topological information between $z^1(t)$ and $z^2(t)$. Since the embeddings in both $z^1(t)$ and $z^2(t)$ can serve as anchor samples, we also define

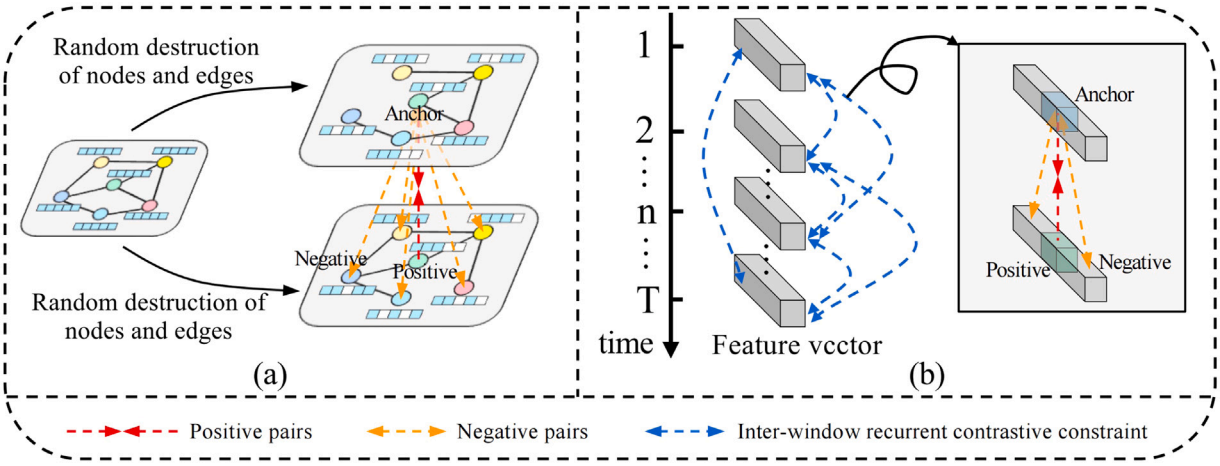


Fig. 3. Illustration of time windows-based dual contrastive constraint. (a) Intra-window contrastive constraint focuses on the discriminative features within a single time window. (b) Inter-window recurrent contrastive constraint captures higher-order temporal dependencies across consecutive time windows.

a contrastive optimization objective as $\eta\left(q_i, p_i\right)$. The intra-window contrastive constraint is shown in Fig. 3(a), and can be expressed as:

$$\mathcal{L}_{intra} = \frac{1}{2N}\sum_{i=1}^{N}\left[\eta(p_i, q_i) + \eta(q_i, p_i)\right] \tag{14}$$

2.3.2. Inter-window recurrent contrastive constraint

In the temporal convolutional contrastive network, we propose an optimization objective $\eta\left(u_i(m), v_i(n)\right)$ to capture higher-order temporal information across windows. The activities of different brain regions interact through complex temporal dependencies, forming intricate network connections. To thoroughly explore the dynamic changes of the brain's temporal sequence, we apply a cross-window recurrent contrastive constraint to the feature embedding S. The inter-window recurrent contrastive constraint is shown in Fig. 3(b), and can be defined as:

$$\mathcal{L}_{inter} = \frac{1}{NT(T-1)}\sum_{m=1}^{T}\sum_{n=1,n\neq m}^{T}\sum_{i=1}^{N}\left(\eta(u_i(m), v_i(n))\right) \tag{15}$$

2.3.3. Loss function

Additionally, we use cross-entropy to compute the classification loss for subjects, denoted as $\mathcal{L}_{CE}$. In order to combine the classification loss with the contrastive constraints, we define the time windows-based dual contrastive loss function as follows:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda_1\mathcal{L}_{intra} + \lambda_2\mathcal{L}_{inter} \tag{16}$$

where $\mathcal{L}_{CE}$ is the cross-entropy loss, and λ_1, λ_2 are hyperparameters that control the relative significance of intra-window discriminative topological information and inter-window higher-order temporal information.

3. Experimental results and analysis

3.1. Data acquisition and preprocessing

In this study, we perform experiments using the epilepsy and the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The details of these datasets are outlined below:

3.1.1. Epilepsy

The rs-fMRI and DTI data for the epilepsy were sourced from Jinling Hospital, which is affiliated with the School of Medicine at Nanjing University. The dataset includes 114 healthy controls (NC) and 192 individuals with epilepsy, where 103 individuals have frontal lobe epilepsy (FLE) and 89 individuals have temporal lobe epilepsy (TLE). Data was acquired from a 3T Siemens Trio MRI scanner. The rs-fMRI images were acquired with the following parameters: voxel size = $3.75 \times 3.75 \times 3.75$ mm^3, TR = 2000 ms, TE = 30 ms, and flip angle = 90°. And voxel size of $0.94 \times 0.94 \times 3$ mm^3, TE = 93 ms, TR = 6100 ms, and flip angle = 90° were used for acquiring DTI scans.

3.1.2. ADNI

In this work, the fMRI and DTI data of 227 subjects from ADNI2 and ADNI3 were analyzed. The subjects were divided into three major categories: 76 individuals showing significant memory concerns (SMC), 69 participants with early mild cognitive impairment (EMCI), and 82 normal controls (NC). Among the NC group, there were 42 females and 40 males, with a mean age of 74 years. The SMC group comprises of 29 females and 47 males, with an average age of 77 years. The EMCI includes 32 females and 37 males, with a mean age of 74 years.

3.1.3. Preprocessing

All preprocessing steps were standardized to provide consistency between the epilepsy and ADNI datasets. Rs-fMRI data were preprocessed using SPM8 and the DPARSF toolbox. First, to minimize the effect of head motion during scanning, slice-timing correction was performed to realign acquisition times across slices. This was followed by realignment to a reference volume in order to improve spatial consistency across all images. The functional images were then aligned to the Montreal Neurological Institute template and resampled to a uniform voxel size of $3 \times 3 \times 3$ mm^3. Linear detrending was then performed, and a bandpass filter removed low-frequency drift and high-frequency noise arising from physiological fluctuations. To further reduce the confounding effects, the motion parameters, global mean signals, and signals from WM and CSF were regressed out as interference covariates to minimize the influence of head movement and non-neuronal fluctuations. After the above preprocessing, segmentation of the rs-fMRI data was carried out by using the Automated Anatomical Labeling (AAL) (Tzourio-Mazoyer et al., 2002) atlas. In the epilepsy dataset, the images were divided into 90 ROIs with 240 time points in each. Similarly, rs-fMRI data from the ADNI dataset was segmented into 90 ROIs with each region consisting of 197 time points. For DTI data, preprocessing was conducted using the PANDA toolbox. Distortions of the DTI images were corrected using the FSL toolbox, and fiber tractography was generated using TrackVis. AAL atlas delineated the anatomical regions according to each subject's T1-weighted images. Fiber count quantification would be the last step in assessing structural connectivity between different regions of the brain.

Table 1
The classification results (%) of the proposed method compared to other approaches on the epilepsy dataset.

Method	NC vs. FLE			NC vs. TLE			NC vs. FLE&TLE			FLE vs. TLE		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
CCA (Weenink, 2003)	69.55 ± 14.56	64.99 ± 20.12	74.04 ± 19.27	72.88 ± 7.82	65.90 ± 10.61	77.26 ± 11.53	73.53 ± 4.24	79.51 ± 3.41	77.20 ± 7.53	60.34 ± 8.42	63.53 ± 9.43	62.29 ± 11.58
DCCA (Andrew et al., 2013)	70.50 ± 6.87	56.52 ± 22.64	65.18 ± 7.94	74.50 ± 11.72	64.09 ± 24.36	68.83 ± 17.23	72.09 ± 3.71	78.65 ± 5.81	59.49 ± 10.49	69.23 ± 8.43	65.46 ± 17.11	62.91 ± 14.77
MKL (Gönen and Alpaydın, 2011)	59.50 ± 11.87	57.84 ± 12.62	60.92 ± 12.94	62.95 ± 10.68	57.44 ± 11.51	63.51 ± 14.27	63.95 ± 8.73	71.92 ± 7.56	60.87 ± 10.14	55.65 ± 9.26	59.82 ± 7.88	46.13 ± 12.49
BrainNetCNN (Kawahara et al., 2017)	73.29 ± 10.31	66.95 ± 14.97	71.59 ± 10.56	73.48 ± 7.71	63.82 ± 11.83	67.69 ± 6.73	72.94 ± 9.27	77.40 ± 9.58	70.39 ± 10.65	69.87 ± 6.91	68.43 ± 17.39	59.36 ± 13.85
SAGPool (Lee et al., 2019)	64.55 ± 7.09	54.93 ± 22.41	61.40 ± 13.47	67.45 ± 8.46	53.05 ± 23.75	59.62 ± 13.14	69.26 ± 7.11	76.11 ± 8.65	59.15 ± 8.38	69.32 ± 9.53	71.20 ± 12.44	65.07 ± 19.51
GCN (Kipf and Welling, 2016)	71.06 ± 11.25	64.82 ± 16.67	72.15 ± 14.59	76.33 ± 7.53	67.88 ± 12.43	72.03 ± 9.05	74.77 ± 6.52	80.40 ± 6.10	71.49 ± 10.78	69.76 ± 5.77	71.33 ± 9.93	67.48 ± 14.15
GAT (Velickovic et al., 2017)	74.18 ± 5.59	69.56 ± 6.08	69.26 ± 9.39	76.90 ± 8.87	69.14 ± 12.00	75.78 ± 15.86	80.70 ± 3.47	85.47 ± 3.22	76.90 ± 4.95	72.42 ± 6.90	74.49 ± 7.21	70.28 ± 10.54
AM-GCN (Wang et al., 2020)	70.06 ± 6.49	54.42 ± 28.84	59.61 ± 14.99	72.88 ± 6.72	57.23 ± 21.76	64.18 ± 10.24	75.86 ± 7.11	81.12 ± 6.93	68.80 ± 12.33	67.74 ± 8.14	66.01 ± 22.78	61.06 ± 14.50
LSTM (Hochreiter, 1997)	72.77 ± 7.99	61.32 ± 23.38	68.75 ± 13.90	73.74 ± 7.82	62.80 ± 11.22	74.54 ± 7.31	77.43 ± 8.10	79.18 ± 16.45	73.25 ± 9.26	66.71 ± 9.84	56.31 ± 27.94	58.12 ± 11.47
ST-GCN (Gadgil et al., 2020)	75.11 ± 6.18	70.28 ± 12.20	70.45 ± 12.50	73.79 ± 4.26	67.54 ± 12.35	72.72 ± 7.49	75.12 ± 6.51	80.50 ± 7.05	74.57 ± 8.21	68.76 ± 6.11	68.64 ± 15.00	66.71 ± 12.58
Ours	**81.49 ± 9.80**	**78.17 ± 9.41**	**79.08 ± 10.00**	**83.50 ± 7.09**	**80.11 ± 9.52**	**80.61 ± 10.20**	**82.67 ± 8.79**	**86.67 ± 6.95**	**80.77 ± 10.90**	**77.22 ± 6.78**	**77.66 ± 7.52**	**80.15 ± 7.96**

Table 2
The classification results (%) of the proposed method compared to other approaches on the ADNI dataset.

Method	NC vs. SMC			NC vs. EMCI			NC vs. SMC&EMCI			SMC vs. EMCI		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
CCA (Weenink, 2003)	83.19 ± 10.87	81.43 ± 12.78	91.05 ± 8.30	81.04 ± 8.08	76.21 ± 10.24	92.71 ± 5.94	82.79 ± 6.94	86.19 ± 6.18	91.10	77.69 ± 13.52	79.03 ± 12.67	82.86 ± 10.74
DCCA (Andrew et al., 2013)	86.92 ± 9.76	86.02 ± 10.28	88.98 ± 11.73	80.77 ± 13.43	73.04 ± 21.66	78.22 ± 18.22	86.08 ± 9.43	87.55 ± 9.89	86.43 ± 11.65	76.92 ± 12.40	77.58 ± 16.10	71.10 ± 19.80
MKL (Gönen and Alpaydın, 2011)	82.43 ± 7.38	78.67 ± 10.97	76.71 ± 15.56	84.72 ± 10.32	79.03 ± 16.68	90.80 ± 9.02	81.31 ± 9.35	84.72 ± 7.83	88.23 ± 7.73	73.07 ± 11.01	71.45 ± 14.95	75.01 ± 13.93
BrainNetCNN (Kawahara et al., 2017)	85.38 ± 8.38	85.06 ± 9.59	84.73 ± 13.21	83.41 ± 6.68	80.45 ± 9.13	82.22 ± 13.99	83.21 ± 7.66	86.64 ± 6.27	87.00 ± 5.97	77.69 ± 11.12	72.46 ± 26.87	76.16 ± 13.84
SAGPool (Lee et al., 2019)	82.29 ± 22.23	83.95 ± 18.13	87.58 ± 15.27	84.51 ± 12.85	82.53 ± 10.33	85.12 ± 14.09	84.33 ± 10.64	88.50 ± 7.63	85.74 ± 13.79	74.62 ± 8.46	64.51 ± 27.70	63.05 ± 17.35
GCN (Kipf and Welling, 2016)	82.29 ± 10.80	82.55 ± 10.91	84.27 ± 12.78	85.71 ± 7.13	81.61 ± 11.78	83.51 ± 12.07	84.69 ± 4.80	87.97 ± 4.25	86.21 ± 5.56	76.92 ± 6.88	76.59 ± 13.62	72.03±10.72
GAT (Velickovic et al., 2017)	86.76 ± 9.43	86.37 ± 9.61	87.63 ± 10.55	86.54 ± 7.88	80.89 ± 13.40	81.38 ± 13.83	88.10 ± 7.51	90.44 ± 6.64	88.99 ± 11.99	80.00 ± 10.99	81.52 ± 10.96	78.55 ± 17.38
AM-GCN (Wang et al., 2020)	88.71 ± 7.95	87.88 ± 6.74	89.71 ± 5.46	86.32 ± 8.99	82.48 ± 12.45	87.33 ± 10.74	86.31 ± 8.27	89.07 ± 7.11	87.49 ± 9.36	83.08 ± 10.77	84.82 ± 10.00	82.70 ± 14.36
LSTM (Hochreiter, 1997)	80.48 ± 12.50	78.89 ± 13.55	81.27 ± 14.74	81.04 ± 6.45	70.85 ± 24.36	78.81 ± 12.01	86.24 ± 6.14	88.15 ± 7.83	85.97 ± 6.30	72.31 ± 12.02	68.89 ± 24.72	73.45 ± 13.44
ST-GCN (Gadgil et al., 2020)	81.76 ± 5.81	71.08 ± 25.57	81.87 ± 15.36	85.88 ± 8.18	79.59 ± 18.19	86.49 ± 11.14	86.21 ± 6.07	88.96 ± 5.68	88.03 ± 7.99	75.38 ± 8.97	77.12 ± 8.57	70.36 ± 8.35
Ours	**92.00 ± 5.99**	**91.71 ± 7.30**	**94.31 ± 6.86**	**94.00 ± 7.99**	**93.50 ± 8.75**	**95.05 ± 11.83**	**94.00 ± 4.67**	**94.62 ± 5.03**	**92.95 ± 8.42**	**88.00 ± 7.48**	**87.85 ± 7.57**	**88.91 ± 9.11**

3.2. Experimental settings

The proposed method is developed using PyTorch and executed on an NVIDIA GeForce RTX 3080 GPU. For each channel of multi-channel GAT, we initially employ two-layer GAT to extract topological features. Then, two-layer MLP is used as projection layer to Decrease the feature dimensionality. Subsequently, the self-attention mechanism is used to integrate features from different channels, providing a comprehensive representation for the brain network. The LSTM deep network is constructed with a two-layer stacked LSTM unit. Finally, an MLP predicts each individual's category through a fully connected layer composed of 128, 32, and 2 neurons. A dropout rate of 0.5 is applied to reduce the risk of overfitting. Optimization is performed using the Adam algorithm, with a learning rate of 2e-4 and weight decay set to 2e-5. The model is trained for 300 epochs, with a batch size of 15.

To assess the effectiveness of our method and other the comparison methods, the preprocessing of fMRI and DTI data is conducted independently for each subject, without any interaction between subjects, ensuring subject independence throughout the preprocessing stage. We conduct experiments on the epilepsy and ADNI datasets using ten-fold cross-validation. The datasets are randomly divided into ten nearly equal, non-overlapping subsets. In each fold, nine subsets are used for training, while the remaining one is used for testing. During the training phase, 20% of the training data is held out for validation. The model parameters that yield the best performance on the validation set are then applied to assess the test set. This procedure is repeated ten times, and we report the average performance and standard deviation across these trials. The experiment setting ensures that the testing set remains completely unseen during both the training and validation phases, preventing any risk of data leakage. To optimize the model, we perform a grid search to determine the best hyperparameters, including the selection of key model parameters such as window sizes T, the probability p_{f1}, p_{f2}, p_{a1}, p_{a2}, and hyperparameters λ_1 and λ_2. Specifically, T is selected from $\{2, 3, 4, 5, 6, 7, 8, 9, 10\}$, p_f, p_a are from $\{0.5, 0.4, 0.3, 0.2, 0.1\}$, and λ_1, λ_2 are from {1e-4, 1e-3, 1e-2, 1e-1, 1, 10}. These hyperparameter combinations are explored in the outer loop of the ten-fold cross-validation to identify the best parameter set. We evaluate the performance of our method on four binary classification tasks from the epilepsy dataset: NC vs. FLE, NC vs. TLE, NC vs. FLE&TLE, and FLE vs. TLE. In a similar manner, the method's performance is also tested on four binary classification tasks in the ADNI dataset: NC vs. SMC, NC vs. EMCI, NC vs. SMC&EMCI, and SMC vs. EMCI. Each classification task is measured using F-Measure (F1), the area under the receiver operating characteristic curve (AUC), and accuracy (ACC).

3.3. Comparison methods

To evaluate the effectiveness of the proposed method, we compare it with ten brain networks analysis methods, including three traditional machine learning methods and seven deep learning-based methods for extracting topological and spatio-temporal features of brain networks. The traditional machine learning methods include canonical correlation analysis (CCA) (Weenink, 2003), deep canonical correlation analysis (DCCA) (Andrew et al., 2013), and multiple kernel learning (MKL) (Gönen and Alpaydın, 2011). The deep learning-based methods consist of brain networks convolutional neural networks (BrainNetCNN) (Kawahara et al., 2017), self-attention graph pooling (SAGPool) (Lee et al., 2019), graph convolutional networks (GCN) (Kipf and Welling, 2016), graph attention networks (GAT) (Velickovic et al., 2017), adaptive multi-channel graph convolutional networks (AM-GCN) (Wang et al., 2020), long short-term memory (LSTM) (Hochreiter, 1997), and spatio-temporal graph convolution networks (ST-GCN) (Gadgil et al., 2020).

(1) CCA: In this work, CCA extracts features of linear correlation after the concatenation of fMRI and DTI data and hence is able to effectively capture the fundamental relationship across multimodal data.

(2) DCCA: In this work, the nonlinear feature transformations are applied to concatenated fMRI and DTI data, allowing for the extraction of complex cross-modal relationships and enabling deeper exploration of implicit patterns between brain structure and function.

(3) MKL: A combination of multiple kernel functions is used for nonlinear feature extraction from concatenated data, flexibly capturing high-dimensional features from multi-modal data and achieving more profound modal fusion.

(4) BrainNetCNN: Specifically designed for DTI datasets, this method introduces three specialized convolutional layer types that aim at leveraging the intrinsic structure of the brain network with weights.

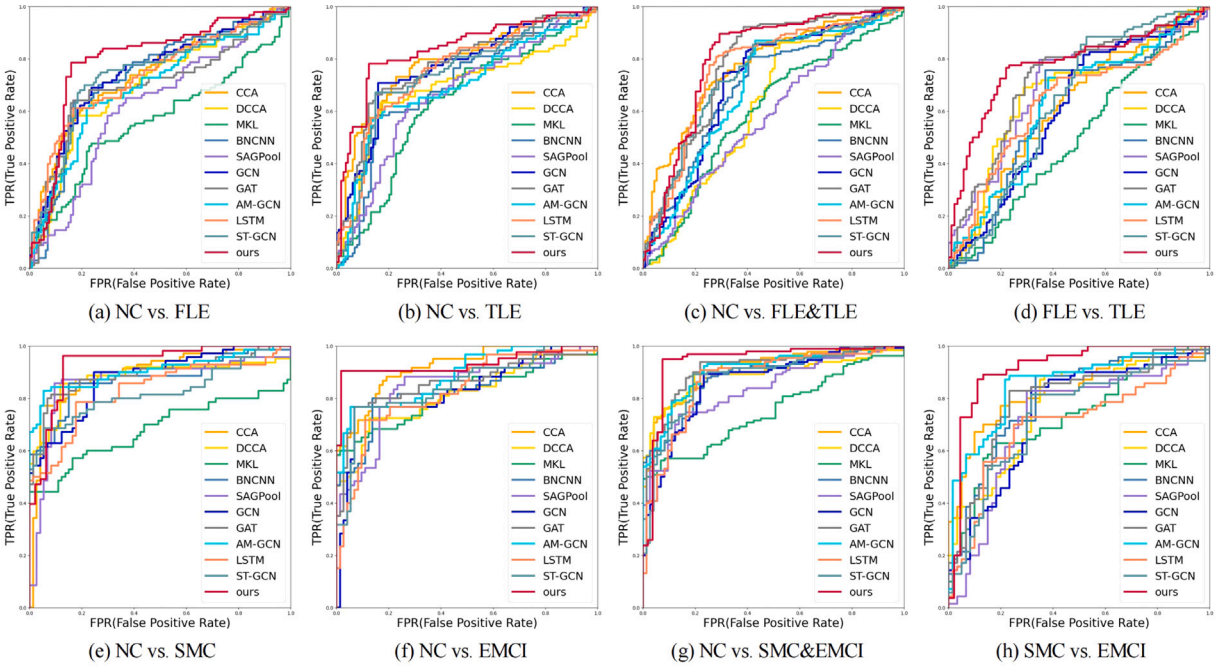


Fig. 4. ROC curves illustrating the performance of the proposed method and comparison approaches for diagnosing brain diseases on the epilepsy and ADNI datasets.

(5) SAGPool: In this method, based on a self-attention pooling mechanism, important nodes are selected to retain essential information, effectively extracting node features and preserving the overall topology of brain networks.

(6) GCN: In GCN, neighboring node features are aggregated layer by layer to capture local structural relationships, revealing interaction patterns between nodes in brain networks and making it suitable for connectivity pattern analysis.

(7) GAT: In GAT, adaptive weights are assigned to node relationships, allowing this method to flexibly capture complex dependencies between nodes and enhance multimodal fusion between fMRI and DTI.

(8) AM-GCN: In this method, self-attention mechanisms are combined with GCN to deeply integrate functional and structural features of brain networks, demonstrating strong performance in joint modeling of multimodal data.

(9) LSTM: In this method, by dividing fMRI data into dynamic time series and combining it with DTI features, it captures the temporal dynamics of brain networks, making it suitable for analyzing temporal variation patterns.

(10) ST-GCN: In ST-GCN, joint temporal and spatial features are extracted to simultaneously capture the temporal dynamics of fMRI and structural dependencies from DTI.

3.4. Classification performance

Tables 1 and 2 present the experimental results of our proposed method and comparison approaches on the epilepsy and ADNI datasets. The data presented show that our method delivers the most accurate diagnostic performance across both datasets. Additionally, the results shown in Fig. 4 reveal that the ROC curve for our method occupies the top-left corner, indicating superior performance in terms of AUC when compared to the other methods. These findings provide strong evidence for the effectiveness of our method in diagnosing brain diseases. Compared to CCA, DCCA, MKL, and BrainNetCNN, our method more effectively preserves the topological structure of brain networks. In contrast to SAGPool, GCN, GAT, and AM-GCN, our method integrates contrastive learning, enhancing the extraction of discriminative topological features. Additionally, in comparison to LSTM and ST-GCN, our method captures higher-order temporal information across time windows, offering a more detailed understanding of temporal dynamics. Our method outperforms others because it not only embeds structural connectivity into functional connectivity but also considers higher-order spatio-temporal feature interactions within and across time windows.

3.5. Ablation study

To verify the performances of the three components of the proposed method, including multi-channel spatial attention contrastive network (MSAC), intra-window contrastive constraint ($\mathcal{L}_{intra}$), and inter-window recurrent contrastive constraint ($\mathcal{L}_{inter}$), we conduct the ablation studies on two datasets, with the results shown in Table 3. The baseline represents brain networks classification only using an MLP. The ✓indicates that the component is used; otherwise, it is not included. After removing MSAC, we replace it with a multi-channel graph convolution network. Schemes 1–6 represent different combinations of components used in the modular ablation experiments to evaluate the contributions of the modules in our proposed method. When comparing Scheme 1, Scheme 2, and Scheme 3 with the baseline, we observe that each component contributes to improving disease diagnosis performance. Additionally, when comparing Scheme 4, Scheme 5, and Scheme 6 with Scheme 1, Scheme 2, and Scheme 3, we find that the results of combining any two components provide better performance compared to using only one. The results also demonstrates that integrating three components can enhance the classification performance. The MSAC incorporates contrastive learning to enhance the extraction of comprehensive topological features. The time windows-based dual contrastive constraint considers dynamic brain network interactions across time windows, enabling the capture of higher-order spatio-temporal information.

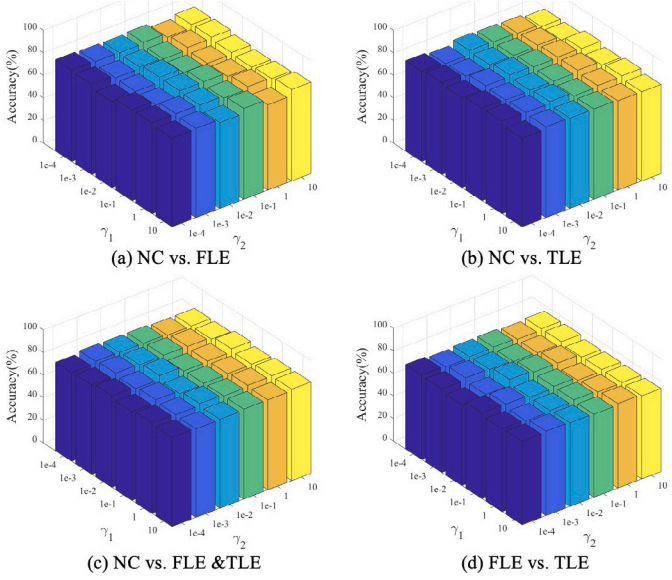
To demonstrate the effectiveness of modality fusion, we evaluate the impact of combining multi-modal data for epilepsy diagnosis. Different modalities, such as fMRI and DTI, provide complementary information that enhances the extraction of disease-discriminative features. As shown in Table 4, using either fMRI or DTI alone achieves good performance. However, fusing both modalities significantly improves classification accuracy. This demonstrates that multi-modal fusion captures complementary features effectively, leading to superior diagnostic performance and highlighting the robustness of the proposed method.

Table 3
The classification performance of the proposed MSTGAC method in modular ablation experiments (%).

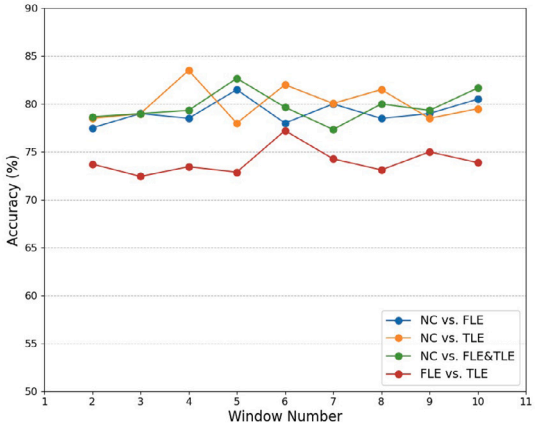
Dataset	Method	MSAC	$\mathcal{L}_{intra}$	$\mathcal{L}_{inter}$	NC vs. FLE			NC vs. TLE			NC vs. FLE&TLE			FLE vs. TLE		
					ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
epilepsy	baseline				71.00	71.11	67.69	70.00	65.16	68.27	71.53	76.63	63.87	69.50	63.32	65.79
	Scheme1	✓			74.00	70.79	71.78	74.50	65.96	69.99	74.00	78.56	70.15	73.33	71.00	64.60
	Scheme2		✓		72.50	69.55	68.74	71.33	64.17	67.10	73.84	76.42	65.16	72.30	62.25	67.67
	Scheme3			✓	73.00	65.50	70.33	72.50	66.80	67.82	72.67	79.09	70.61	73.07	69.93	68.87
	Scheme4	✓	✓		78.49	72.53	70.61	77.50	74.14	76.45	78.67	83.57	69.60	74.67	77.21	63.62
	Scheme5	✓		✓	79.50	67.46	75.63	79.99	61.11	75.19	78.99	82.95	69.29	76.50	72.57	67.95
	Scheme6		✓	✓	75.50	71.13	76.40	76.00	65.91	67.11	74.50	60.69	70.50	73.50	60.29	74.19
	Ours	✓	✓	✓	**81.49**	**78.17**	**79.08**	**83.50**	**80.11**	**80.61**	**82.67**	**86.67**	**80.77**	**77.22**	**77.66**	**80.15**
					NC vs. SMC			NC vs. EMCI			NC vs. SMC&EMCI			SMC vs. EMCI		
ADNI	baseline				82.00	71.55	61.61	80.99	68.01	67.87	82.22	79.23	80.61	80.00	67.77	76.77
	Scheme1	✓			85.83	85.08	86.67	86.66	75.42	92.54	83.50	83.97	88.13	84.01	83.74	84.63
	Scheme2		✓		80.00	77.97	75.64	82.22	75.30	67.53	82.67	84.52	80.78	80.99	84.54	78.94
	Scheme3			✓	83.01	78.43	80.99	81.67	78.74	81.29	84.67	81.25	76.80	83.00	75.43	63.75
	Scheme4	✓	✓		89.00	86.33	84.35	87.50	85.85	87.87	86.67	87.19	87.04	87.77	85.83	79.38
	Scheme5	✓		✓	87.50	89.54	86.04	91.00	82.85	83.04	89.00	89.94	89.96	86.01	85.39	87.57
	Scheme6		✓	✓	87.00	72.77	81.49	86.01	80.74	83.72	88.50	89.02	88.76	85.00	78.08	75.65
	Ours	✓	✓	✓	**92.00**	**91.71**	**94.31**	**94.00**	**93.50**	**95.05**	**94.00**	**94.62**	**92.95**	**88.00**	**87.85**	**88.91**

Table 4
The classification performance of the proposed MSTGAC method in multi-modal ablation experiments (%).

Dataset	Method	fMRI	DTI	NC vs. FLE			NC vs. TLE			NC vs. FLE&TLE			FLE vs. TLE		
				ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
epilepsy	Scheme1	✓		77.99	65.25	72.83	79.49	72.83	76.20	78.67	83.30	77.36	74.17	70.88	62.42
	Scheme2		✓	79.23	58.49	70.58	78.67	62.91	74.70	74.33	79.99	71.23	75.33	66.75	74.02
	Ours	✓	✓	**81.49**	**78.17**	**79.08**	**83.50**	**80.11**	**80.61**	**82.67**	**86.67**	**80.77**	**77.22**	**77.66**	**80.15**
				NC vs. SMC			NC vs. EMCI			NC vs. SMC&EMCI			SMC vs. EMCI		
ADNI	Scheme1	✓		85.56	79.57	87.33	87.00	75.34	88.45	91.33	89.47	91.64	85.00	78.08	75.65
	Scheme2		✓	86.00	81.26	85.89	85.00	72.46	80.47	87.99	90.69	87.56	84.17	84.54	83.89
	Ours	✓	✓	**92.00**	**91.71**	**94.31**	**94.00**	**93.50**	**95.05**	**94.00**	**94.62**	**92.95**	**88.00**	**87.85**	**88.91**



(a) The accuracy (%) of our method under different parameter combinations across four diagnostic tasks.



(b) The impact of varying window numbers on the model's classification performance.

Fig. 5. Impact of parameter combinations and window numbers on classification accuracy.

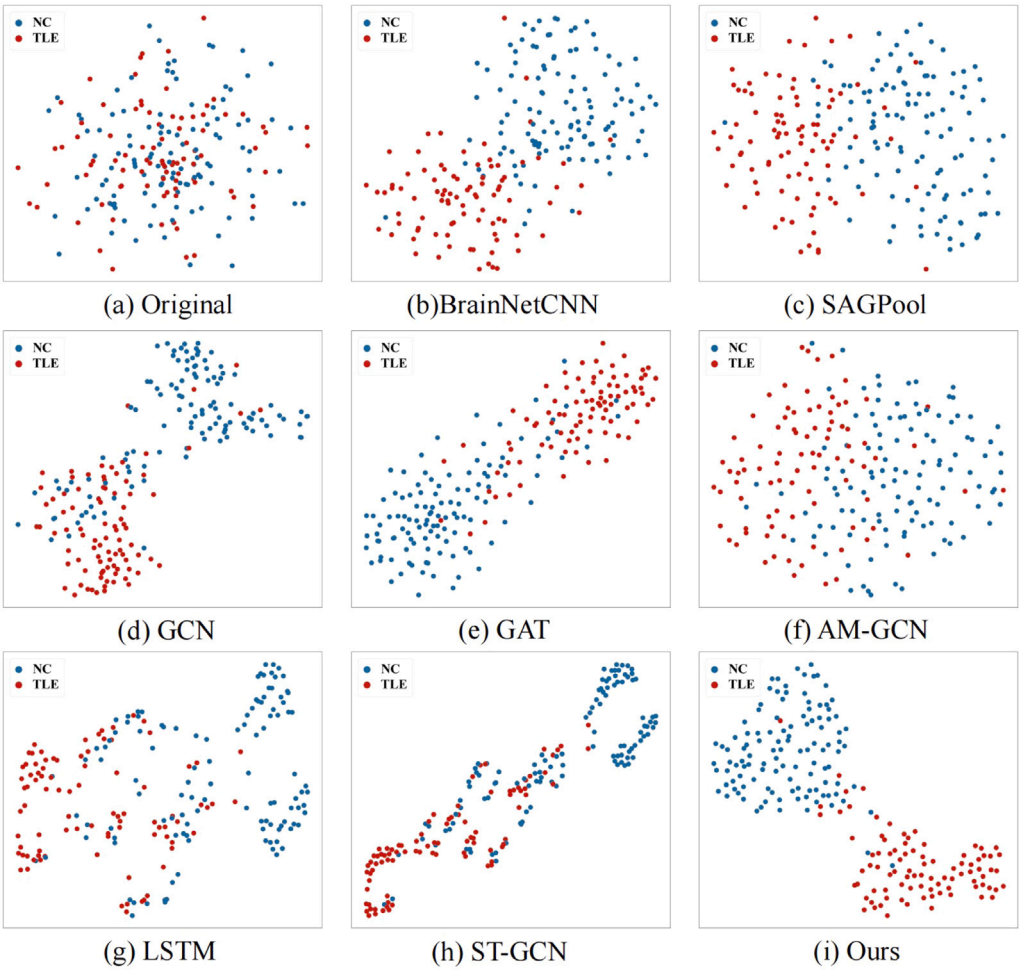


Fig. 6. The t-SNE plots showcasing the performance of various methods for the NC vs. TLE task on the epilepsy dataset.

4. Discussion

4.1. Analysis of parameter sensitivity

As indicated in Eq. (16), the total loss function involves two parameters. To assess the impact of these parameters on our experimental outcomes and to identify the optimal values, we utilize a grid search technique. Specifically, the parameters λ_1 and λ_2 are chosen from the set $\{1e\text{-}4, 1e\text{-}3, 1e\text{-}2, 1e\text{-}1, 1, 10\}$. The results of this parameter analysis are shown in Fig. 5(a). Overall, it is evident from Fig. 5 that our model performs consistently well across the four epilepsy-related tasks, regardless of the specific values chosen for λ_1 and λ_2. This suggests that the proposed method is robust with respect to these parameters. Additionally, we explore the effect of the number of time windows on the experimental results. Specifically, we construct DBNs with varying numbers of sliding windows T selected from $\{2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and evaluate performance across four classification tasks. As shown in Fig. 5(b), for the NC vs. TLE task, classification accuracy peaks when the number of windows is 4. For the FLE vs. TLE task, the highest accuracy is achieved when $T = 6$. Similarly, the best performance for both the NC vs. FLE task and the NC vs. FLE&TLE task is obtained with $T = 5$. Moreover, we observe a consistent trend across all four tasks: as the number of sliding windows increases, classification accuracy first improves, reaches a maximum, and then starts to decline. This trend is justifiable because with too few windows, the interactions between ROIs across different time windows may be insufficient, leading to information loss. On the other hand, when the number of windows becomes too large, excessive partitioning may introduce redundant information within each window, thereby reducing classification accuracy.

4.2. Analysis of feature embeddings

To further verify the ability of our proposed model to extract discriminative spatio-temporal information, we conduct the task of visualization for the NC vs. TLE task on epilepsy dataset. We extract the embeddings from the final layer of MSTGAC before the SoftMax operation for t-SNE analysis (Van der Maaten and Hinton, 2008), which enables the visualization of high-dimensional data in a low-dimensional space. The original sample data are visualized in Fig. 6(a), and the representations of our method and several comparison methods are shown in Fig. 6(b)–(i). Specifically, the original data exhibit a distribution which is difficult to distinguish. Although BrainNetCNN, SAGPool, GCN, GAT, and AM-GCN can capture topological information across modality, they fail to capture temporal information within the networks. The LSTM and ST-GCN can learn complex temporal information in brain networks, but fail to capture higher-order spatio-temporal features. In Fig. 6(i), samples of the same class are closely clustered, while clear boundaries exist between samples of different classes. Our method not only learns discriminative spatio-temporal features but also leverages contrastive learning to minimize the distance between samples from same class while increasing the separation between samples from different classes.

4.3. Discriminative brain regions and connectivity on epilepsy

We assessed the efficacy of the proposed method in identifying higher-order brain connections that are highly discriminative for epil epsy. To investigate the discriminative power of brain regions and their connections, we utilized a non-negative elastic net regression approach (Zhu et al., 2018). Specifically, we constructed a regression model that links the embeddings from the last layer of MSTGAC (prior to the SoftMax operation) to the diagnostic labels using this technique. The regression coefficients obtained from this model represent significant differences in connectivity between healthy control and the patient groups. To identify the most discriminative brain connections, we ranked these connections based on the absolute values of the regression coefficients. We aggregated the regression coefficients of the brain connections across all folds during ten-fold cross-validation to identify the top 20 most discriminative brain connections and visualize them in Fig. 7(a).

For instance, in NC vs. FLE, the significant connections are mostly in regions including supplementary motor area of the frontal lobe, inferior frontal gyrus, and middle frontal gyrus that are closely related to cognitive control and speech expression (Carney et al., 2012; Bonini et al., 2014). Cognitive impairment and speech arrest are widely reported symptoms of FLE, and these brain connections might be impaired in FLE patients. In NC vs. TLE, the most distinguishing connections exist in the hippocampus, inferior temporal gyrus, and middle temporal gyrus regions related to memory formation and language and visual information processing (Herlin et al., 2021; Whitten et al., 2021). Symptoms in patients with TLE include memory impairment and auditory and visual hallucinations. That suggests that altered connectivity in such regions might be responsible for memory impairment and the development of hallucinations frequently observed in TLE patients. In NC vs. FLE&TLE, SACs are concentrated in the superior temporal gyrus, rolandic operculum, and superior frontal gyrus regions associated with auditory and language comprehension (Trimmel et al., 2021; Wang et al., 2021). These regions represent areas of common dysfunction of sensory and language processing for both subtypes of epilepsy. In FLE vs. TLE, SACs are primarily located in critical regions such as supplementary motor area, amygdala, and precentral gyrus, which are critical for motor control (Arbune et al., 2018; Zhu et al., 2021) and emotional processing (Aroniadou-Anderjaska et al., 2008), corresponding to the distinct symptoms observed in FLE and TLE episodes. This highlights the differential influence of FLE and TLE on motor and emotional regulation.

Furthermore, we assessed the discriminative power of each brain region by aggregating the regression coefficients of all connections linked to that region. Subsequently, the brain regions were ranked based on their total regression coefficients, and the top 10 most distinctive regions were identified for each classification task. Each color corresponds to a specific brain region, and the visualization highlights the spatial distribution of these regions across the brain. In Fig. 7(b), in the NC vs. FLE, the important brain regions include the inferior frontal gyrus, middle frontal gyrus, and paracentral lobule. FLE patients may exhibit impairments in language expression and higher cognitive functions, suggesting their significance on epilepsy diagnosis (Bookheimer, 2002; Duncan and Owen, 2000). These brain regions are closely associated with language comprehension and attention control. In the other three tasks, the important regions include temporal pole: middle temporal gyrus, the superior frontal gyrus, olfactory cortex, which have been proven effective in identifying epilepsy and its subtypes (Koechlin et al., 2003; Adcock et al., 2003; Menassa et al., 2017). Moreover, we find several discriminative brain regions outside the frontal and temporal lobes, such as the inferior occipital lobe, superior parietal gyrus, postcentral gyrus, and cerebellar white matter. These regions may be related to the long-term epilepsy activity inducing changes in brain neuroplasticity (Sutula, 2004; Jarero-Basulto et al., 2018). Specifically, the abnormal neural activity caused by epilepsy can result in the significant changes in brain structure and function, extending beyond the area where epilepsy originates to non-focal regions. For example, long-term epileptic activity may promote the formation of new synapses while reducing or altering existing synaptic connections. These changes in synaptic plasticity can affect multiple areas of the brain, including those far from the origin of the epilepsy. The results of these experiments confirm that our approach yields robust biomarkers for the diagnosis of epilepsy.

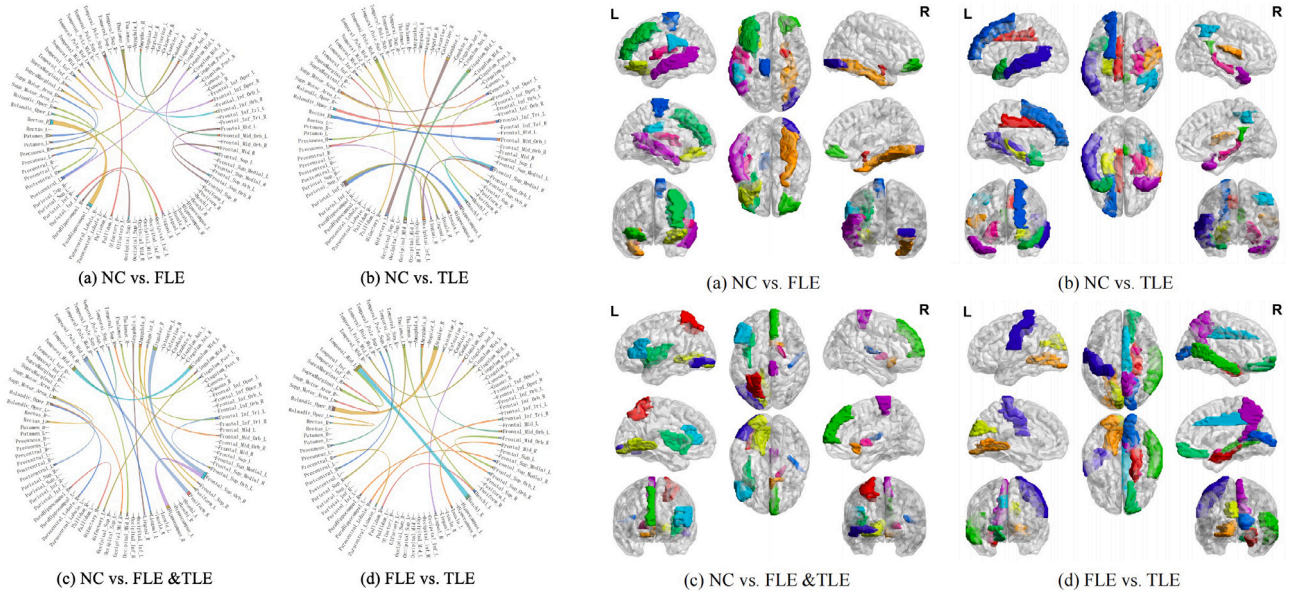
4.4. Impact of global signal regression on identifying discriminative brain regions

This study investigated the impact of global signal regression (GSR) (Xu et al., 2018; Li et al., 2024) on identifying discriminative brain regions for classifying NC vs. SMC&EMCI on the ADNI dataset. The application of GSR introduced differences in discriminative power of brain regions, emphasizing its influence on functional connectivity patterns and biomarker discovery.

When GSR was applied, as shown in Fig. 8(a), the top discriminative regions included the left hippocampus (HIP. L), right hippocampus (HIP. R), left middle temporal gyrus (MTG. L), left amygdala (AMY. L), left precuneus (PRE. L), right angular gyrus (ANG. R), right temporal pole: middle temporal gyrus (TEM. R), left inferior frontal gyrus, triangular part (IFG. L), left parahippocampal gyrus (PAR. L), and right middle frontal gyrus (MFG. R). Several regions are strongly aligned with the established neuropathological features of Alzheimer's disease, particularly in its early stages. The hippocampus and parahippocampal gyrus are central to episodic memory and contextual processing (Smith, 2008), while the precuneus, a critical node in the default mode network (DMN), is involved in memory retrieval and self-referential tasks (Sajonz et al., 2010). Temporal and frontal regions, such as the middle temporal gyrus and angular gyrus, further support semantic memory and language comprehension (Davey et al., 2016). GSR appears to enhance the discriminative power of these regions by mitigating global noise and amplifying localized neural variability critical for detecting disease-related alterations.

In the absence of GSR, as shown in Fig. 8(b), the top discriminative regions shifted significantly, with the identified regions including the right hippocampus (HIP. R), left amygdala (AMY. L), left inferior parietal lobule (INF. L), left inferior frontal gyrus, orbital part (IFG. L), right superior frontal gyrus, medial orbital (SFG. R), right middle frontal gyrus, orbital part (MFG. R), right posterior cingulate gyrus (PCG. R), right postcentral gyrus (POS. R), left middle frontal gyrus (MFG. L), and right lenticular nucleus, putamen (LNP. R). While some overlap with the GSR condition was observed (e.g., the hippocampus and amygdala), the prominence of regions like the inferior parietal lobule and posterior cingulate gyrus suggests a stronger emphasis on global network-level dynamics without GSR. These regions are involved in attention allocation, cognitive integration, and systemic interactions within the DMN (Naghavi and Nyberg, 2005; Leech and Sharp, 2014). However, core Alzheimer's-related regions, such as the precuneus and middle temporal gyrus, exhibited reduced discriminative power, potentially due to the retention of global fluctuations diluting localized disease-relevant signals.

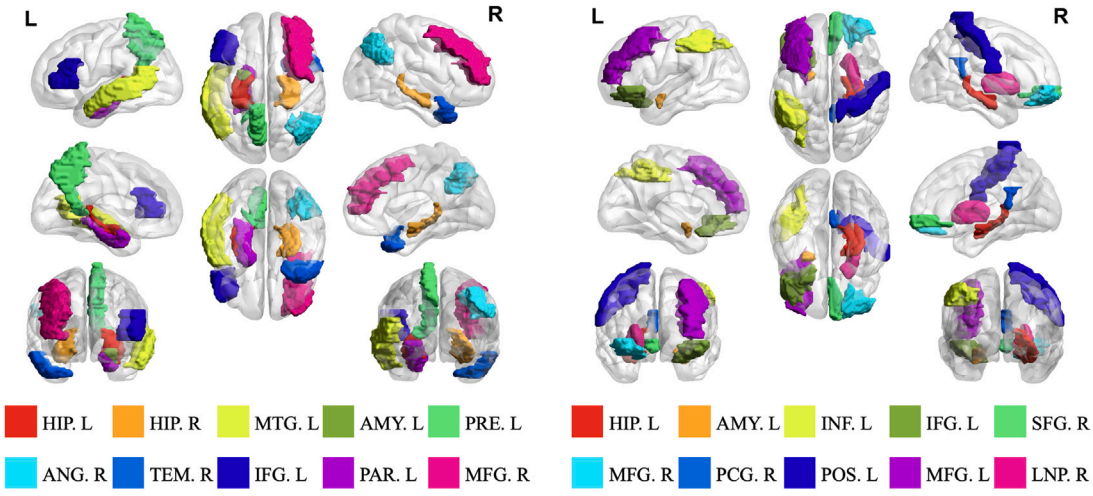
GSR modifies the functional connectivity landscape by removing global signal fluctuations, enhancing the signal-to-noise ratio of localized neural activity. This improves the sensitivity to regions directly implicated in Alzheimer's pathology, such as the hippocampus and precuneus. However, GSR also introduces negative correlations between brain regions, which may distort network dynamics and influence the representation of global connectivity patterns. Conversely, the absence of GSR retains global fluctuations, preserving systemic connectivity features but potentially attenuating the discriminative power of localized neural activity critical for identifying disease-specific biomarkers. The findings underscore the significant influence of GSR on biomarker identification. By emphasizing localized features, GSR enhances the



(a) The 20 most important brain connections identified across various classification tasks.

(b) The spatial distribution of the top 10 most distinguishing brain regions across different classification tasks.

Fig. 7. Analysis of significant brain connections and regions in classification tasks.



(a)The spatial distribution of the top 10 most distinguishing brain regions identified with GSR on the ADNI dataset.

(b)The spatial distribution of the top 10 most distinguishing brain regions identified without GSR on the ADNI dataset.

Fig. 8. Comparison of the spatial distribution of the top 10 most distinguishing brain regions identified with and without GSR on the ADNI dataset.

detection of regions directly linked to Alzheimer's pathology, such as the hippocampus, parahippocampal gyrus, and middle temporal gyrus. Without GSR, the analysis prioritizes broader network-level interactions, leading to the emergence of regions like the posterior cingulate gyrus and motor-related areas. These results highlight the critical role of preprocessing choices in shaping the interpretation and applicability of biomarkers in functional connectivity studies.

5. Conclusion

In this paper, we propose a novel multi-channel spatio-temporal graph attention contrastive network framework, which can be used to capture the higher-order spatio-temporal topological features in DBNs. The proposed method can fuse the functional and structural connectivity for constructing multi-modal brain networks, which provides a comprehensive view for brain networks analysis. Then, we develop a multi-channel spatial attention contrastive network to effectively extract discriminative features from brain networks. Specifically, we design intra-window and inter-window contrastive learning techniques to improve the model's robustness and enhance its ability to capture spatio-temporal dependency features across time windows. The experiments on several brain disorder datasets indicate that the proposed method can not only outperform the existing brain networks analysis methods in diagnostic accuracy but also effectively identify the critical biomarkers, including specific brain regions and neural connections.

CRediT authorship contribution statement

Chaojun Li: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis. **Kai Ma:** Writing – review & editing. **Shengrong Li:** Writing – review & editing. **Xiangshui Meng:** Data curation. **Ran Wang:** Writing –

review & editing. **Daoqiang Zhang:** Writing – review & editing. **Qi Zhu:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Ethics statement

This study was conducted using data that adhered to ethical guidelines, institutional approvals, and established ethical standards, as described below.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/) adhered to ethical protocols established by the institutional review boards of all participating institutions and the United States Code of Federal Regulations. Approval for the use of this dataset was granted by the ADNI Data and Publications Committee prior to the commencement of this study.

The epilepsy dataset was collected at Jinling Hospital, Nanjing University School of Medicine, and approved by the medical ethics committee in Jinling Hospital, Nanjing University School of Medicine. All epilepsy patients included in this study were diagnosed by experienced neurologists. All procedures involving human participants were conducted in accordance with the principles outlined in the Declaration of Helsinki, and written informed consent was obtained from all participants or their legal guardians.

Declaration of competing interest

The authors declare that there are no known financial or personal conflicts of interest that could have influenced the research presented in this paper.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (Nos. 62371234, 62076129, 62136004, 62272226 and 62276130), Natural Science Foundation of Jiangsu Province (No. BK20231438), Key Research and Development Plan of Jiangsu Province (No. BE2022842), and also National Key R&D Program of China (No. 2023YFF1204803).

Data and code availability statement

The authors do not have permission to share epilepsy data. The ADNI data is available at http://adni.loni.usc.edu/. The code has been released on https://github.com/xbrainnet/MSTGAC.git.

References

Adcock, J.E., Wise, R.G., Oxbury, J., Oxbury, S., Matthews, P.M., 2003. Quantitative fMRI assessment of the differences in lateralization of language-related brain activation in patients with temporal lobe epilepsy. Neuroimage 18 (2), 423–438.

Andrew, G., Arora, R., Bilmes, J., Livescu, K., 2013. Deep canonical correlation analysis. In: International Conference on Machine Learning. PMLR, pp. 1247–1255.

Arbune, A.A., Bajenaru, O.A., Popa, I., Mîndruta, I., 2018. Supplementary motor area–clinical semiology and results of direct electrical stimulation during intracranial electrodes exploration. Rom. J. Neurol. 17 (2), 71.

Aroniadou-Anderjaska, V., Fritsch, B., Qashu, F., Braga, M.F., 2008. Pathology and pathophysiology of the amygdala in epileptogenesis and epilepsy. Epilepsy Res. 78 (2–3), 102–116.

Batista-García-Ramó, K., Fernández-Verdecia, C.I., 2018. What we know about the brain structure–function relationship. Behav. Sci. 8 (4), 39.

Bonini, F., McGonigal, A., Trébuchon, A., Gavaret, M., Bartolomei, F., Giusiano, B., Chauvel, P., 2014. Frontal lobe seizures: from clinical semiology to localization. Epilepsia 55 (2), 264–277.

Bookheimer, S., 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu. Rev. Neurosci. 25 (1), 151–188.

Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Rev. Neurosci. 10 (3), 186–198.

Carney, P.W., Masterton, R.A., Flanagan, D., Berkovic, S.F., Jackson, G.D., 2012. The frontal lobe in absence epilepsy: EEG-fMRI findings. Neurology 78 (15), 1157–1165.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. PMLR, pp. 1597–1607.

Chung, M.K., Das, S., Ombao, H., 2024. Dynamic topological data analysis of functional human brain networks. Found. Data Sci. 6 (1), 22–40.

Davey, J., Thompson, H.E., Hallam, G., Karapanagiotidis, T., Murphy, C., De Caso, I., Krieger-Redwood, K., Bernhardt, B.C., Smallwood, J., Jefferies, E., 2016. Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. Neuroimage 137, 165–177.

Drenthen, G.S., Fasen, F., Wald, E.L.F., Backes, W.H., Aldenkamp, A.P., Vermeulen, R.J., Debeij-van Hall, M., Hendriksen, J., Klinkenberg, S., Jansen, J.F., 2020. Functional brain network characteristics are associated with epilepsy severity in childhood absence epilepsy. NeuroImage: Clin. 27, 102264.

Duncan, J., Owen, A.M., 2000. Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends Neurosci. 23 (10), 475–483.

Gadgil, S., Zhao, Q., Pfefferbaum, A., Sullivan, E.V., Adeli, E., Pohl, K.M., 2020. Spatio-temporal graph convolution for resting-state fMRI analysis. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23. Springer, pp. 528–538.

Gai, Q., Chu, T., Che, K., Li, Y., Dong, F., Zhang, H., Li, Q., Ma, H., Shi, Y., Zhao, F., et al., 2023. Classification of major depressive disorder based on integrated temporal and spatial functional mri variability features of dynamic brain network. J. Magn. Reson. Imaging 58 (3), 827–837.

Gönen, M., Alpaydın, E., 2011. Multiple kernel learning algorithms. J. Mach. Learn. Res. 12, 2211–2268.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. PLoS Biol. 6 (7), e159.

Herlin, B., Navarro, V., Dupont, S., 2021. The temporal pole: From anatomy to function—A literature appraisal. J. Chem. Neuroanat. 113, 101925.

Hochreiter, S., 1997. Long Short-term Memory. Neural Computation MIT-Press.

Honey, C.J., Thivierge, J.-P., Sporns, O., 2010. Can structure predict function in the human brain? Neuroimage 52 (3), 766–776.

Jarero-Basulto, J.J., Gasca-Martínez, Y., Rivera-Cervantes, M.C., Ureña-Guerrero, M.E., Feria-Velasco, A.I., Beas-Zarate, C., 2018. Interactions between epilepsy and plasticity. Pharmaceuticals 11 (1), 17.

Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G., 2017. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage 146, 1038–1049.

Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Koechlin, E., Ody, C., Kouneiher, F., 2003. The architecture of cognitive control in the human prefrontal cortex. Science 302 (5648), 1181–1185.

Kottaram, A., Johnston, L., Ganella, E., Pantelis, C., Kotagiri, R., Zalesky, A., 2018. Spatio-temporal dynamics of resting-state brain networks improve single-subject prediction of schizophrenia diagnosis. Hum. Brain Mapp. 39 (9), 3663–3681.

Lee, J., Lee, I., Kang, J., 2019. Self-attention graph pooling. In: International Conference on Machine Learning. pmlr, pp. 3734–3743.

Leech, R., Sharp, D.J., 2014. The role of the posterior cingulate cortex in cognition and disease. Brain 137 (1), 12–32.

Li, X., Bianchini Esper, N., Ai, L., Giavasis, S., Jin, H., Feczko, E., Xu, T., Clucas, J., Franco, A., Sólon Heinsfeld, A., et al., 2024. Moving beyond processing-and analysis-related variation in resting-state functional brain imaging. Nat. Hum. Behav. 8 (10), 2003–2017.

Liu, J., 2024. Topological analysis of overlapping dynamic brain networks under naturalistic stimuli based on sparse non-negative matrix factorization algorithm. Biomed. Signal Process. Control 87, 105527.

Márquez, F., Yassa, M.A., 2019. Neuroimaging biomarkers for Alzheimer's disease. Mol. Neurodegener. 14 (1), 21.

Menassa, D.A., Sloan, C., Chance, S.A., 2017. Primary olfactory cortex in autism and epilepsy: increased glial cells in autism. Brain Pathol. 27 (4), 437–448.

Naghavi, H.R., Nyberg, L., 2005. Common fronto-parietal activity in attention, memory, and consciousness: shared demands on integration? Conscious. Cogn. 14 (2), 390–425.

Sajonz, B., Kahnt, T., Margulies, D.S., Park, S.Q., Wittmann, A., Stoy, M., Ströhle, A., Heinz, A., Northoff, G., Bermpohl, F., 2010. Delineating self-referential processing from episodic memory retrieval: common and dissociable networks. Neuroimage 50 (4), 1606–1617.

Smith, D.M., 2008. The hippocampus, context processing and episodic memory. Handb. Behav. Neurosci. 18, 465–630.

Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., et al., 2013. Resting-state fMRI in the human connectome project. Neuroimage 80, 144–168.

Sutula, T.P., 2004. Mechanisms of epilepsy progression: current theories and perspectives from neuroplasticity in adulthood and development. Epilepsy Res. 60 (2–3), 161–171.

Trimmel, K., Vos, S.B., Caciagli, L., Xiao, F., van Graan, L.A., Winston, G.P., Koepp, M.J., Thompson, P.J., Duncan, J.S., 2021. Decoupling of functional and structural language networks in temporal lobe epilepsy. Epilepsia 62 (12), 2941–2954.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15 (1), 273–289.

Van Den Heuvel, M.P., Sporns, O., 2011. Rich-club organization of the human connectome. J. Neurosci. 31 (44), 15775–15786.

Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (11).

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al., 2017. Graph attention networks. Stat 1050 (20), 10–48550.

Wang, Y., Wang, X., Zhao, B.t., Sang, L., Zhang, C., Hu, W.h., Shao, X.q., Zhang, J.g., Zhang, K., 2021. The rolandic operculum generates different semiologies in insulo-opercular and temporal lobe epilepsies. Epilepsy & Behav. 114, 107614.

Wang, X., Zhu, M., Bo, D., Cui, P., Shi, C., Pei, J., 2020. Am-gcn: Adaptive multi-channel graph convolutional networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1243–1253.

Weenink, D., 2003. Canonical correlation analysis. In: Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, vol. 25, University of Amsterdam Amsterdam, pp. 81–99.

Whitten, A., Jacobs, M.L., Englot, D.J., Rogers, B.P., Levine, K.K., González, H.F., Morgan, V.L., 2021. Resting-state hippocampal networks related to language processing reveal unique patterns in temporal lobe epilepsy. Epilepsy & Behav. 117, 107834.

Xu, H., Su, J., Qin, J., Li, M., Zeng, L.-L., Hu, D., Shen, H., 2018. Impact of global signal regression on characterizing dynamic functional connectivity and brain states. Neuroimage 173, 127–145.

Yu, Q., A Allen, E., Sui, J., R Arbabshirani, M., Pearlson, G., D Calhoun, V., 2012. Brain connectivity networks in schizophrenia underlying resting state functional magnetic resonance imaging. Curr. Top. Med. Chem. 12 (21), 2415–2425.

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L.L., Breakspear, M., 2014. Time-resolved resting-state brain networks. Proc. Natl. Acad. Sci. 111 (28), 10341–10346.

Zhang, W., Tang, F., Zhou, X., Li, H., 2020. Dynamic reconfiguration of functional topology in human brain networks: from resting to task states. Neural Plast. 2020 (1), 8837615.

Zhu, Q., Huang, J., Xu, X., 2018. Non-negative discriminative brain functional connectivity for identifying schizophrenia on resting-state fMRI. Biomed. Eng. Online 17, 1–15.

Zhu, J., Xu, C., Zhang, X., Qiao, L., Wang, X., Zhang, X., Yan, X., Ni, D., Yu, T., Zhang, G., et al., 2021. The thalamus-precentral gyrus functional connectivity changes in epilepsy patients following vagal nerve stimulation. Neurosci. Lett. 751, 135815.