Multi-Modal Brain Network Fusion for Intelligent Diagnostic Devices

Shengrong Li, Qi Zhu, Liang Sun, Kai Ma, Yixin Ji, Shile Qi, and Daoqiang Zhang, *Senior Member, IEEE*

***Abstract*—Embedding multi-modal brain network analysis technology into consumer electronics, such as smart wearables, helps enable early intelligent diagnosis of brain diseases. Recent studies confirm that the functional-structural coupling in certain regions is more tightly correlated than in others. However, existing multi-modal methods often directly fuse functional brain networks (FBN) and structural brain networks (SBN), ignoring the regional heterogeneity between them. Additionally, identity information encoded in brain networks may interfere with disease diagnosis. In this paper, we develop a multi-modal brain network fusion method with regional heterogeneity constraints, and design a feature decoupling module to alleviate disease-irrelevant information. Specifically, we first divide FBN and SBN into multiple subnetworks, and introduce penalty weights to reduce the communication cost between cross-modal brain regions within the subnetwork while increasing the cost between different subnetworks. Then, under the regional heterogeneity constraints, we adopt optimal transport to simulate the transfer of brain region hubness from FBN to SBN, thereby effectively integrating the complex cross-modal interactions. Furthermore, we design a feature decoupling module to suppress ineffective features and enhance the discrimination between modality-specific features and multi-modal features. Experimental results show that the proposed method has promising performance and can identify multi-modal biomarkers for brain disease diagnosis.**

***Index Terms*—Multi-modal brain network, regional heterogeneity, optimal transport, feature decoupling, brain disease diagnosis.**

I. Introduction

With the rapid development of artificial intelligence and Internet of Things technology, telemedicine has garnered increasing attention [1]. As a critical component of healthcare, brain disease diagnosis also benefits from these cutting-edge technological advancements. Wearable devices and intelligent diagnostic systems can identify brain diseases earlier and more accurately by detecting and analyzing the brain's physiological activities [2]–[4]. Therefore, intelligent brain disease diagnosis possesses substantial application potential in both the healthcare and consumer electronics markets [5].

Brain networks can comprehensively describe the complex interactions and association patterns between brain regions, making them a powerful tool for brain disease diagnosis and cognitive science research [6]. Brain networks can be divided into functional brain network (FBN) and structural brain networks (SBN). The FBN is typically constructed using functional magnetic resonance imaging, while the SBN is constructed using diffusion tensor imaging. They can provide complementary information for brain disease diagnosis from different perspectives. However, due to the complex topological coupling relationship between FBN and SBN, it is still a challenge to develop effective multi-modal brain network fusion methods to improve the diagnostic performance of brain diseases.

Recently, several multi-modal brain network analysis methods have been proposed. As shown in Fig. 1, they are primarily divided into two categories: simple fusion and interactive fusion. The simple fusion strategy is shown in Fig. 1(a), it involves the straightforward concatenation or addition of FBN and SBN, and then extracts the shallow structural features from them. For example, Dai et al. first used multiple kernel learning to integrate FBN and SBN, and then employed support vector machine to distinguish between healthy individuals and those with attention deficit disorders [7]. Sebenius et al. adopted graph convolutional network (GCN) to extract the structural information from both FBN and SBN respectively, and concatenated the features of the most important k nodes from each network for classification [8]. In fact, there is a nonlinear mapping relationship between FBN and SBN, and the simple fusion method may lose rich multi-modal coupling information. In order to model the complex intrinsic dependencies between FBN and SBN, researchers have developed an interactive fusion strategy. The interactive fusion strategy is presented in Fig. 1(b), it refers to adaptively mining complementary information between functional and structural networks, and revealing potential associations between different modalities. For instance, Zhu et al. designed a triplet network combining FBN and SBN, and used the attention mechanism to adaptively extract the most important features in the two modalities [9]. Yang et al. proposed a cross-modal graph neural network to capture the deep dependencies between FBN and SBN through dynamic graph learning and mutual learning [10]. However, FBN and SBN are not a one-to-one perfect relationship, and the direct interactive fusion method may ignore the regional heterogeneity between function and structure in the brain.

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), National Key R&D Program of China (No. 2023YFF1204803), and also Key Research and Development Plan of Jiangsu Province (No. BE2022842). Corresponding author: Qi Zhu (zhuqinuaa@163.com).

Shengrong Li, Qi Zhu, Liang Sun, Yixin Ji, Shile Qi, and Daoqiang Zhang are with the College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China, and with Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, University of Aeronautics and Astronautics, Nanjing, 211106, China.

Kai Ma is with the School of Computer, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu Province, 212100, China.

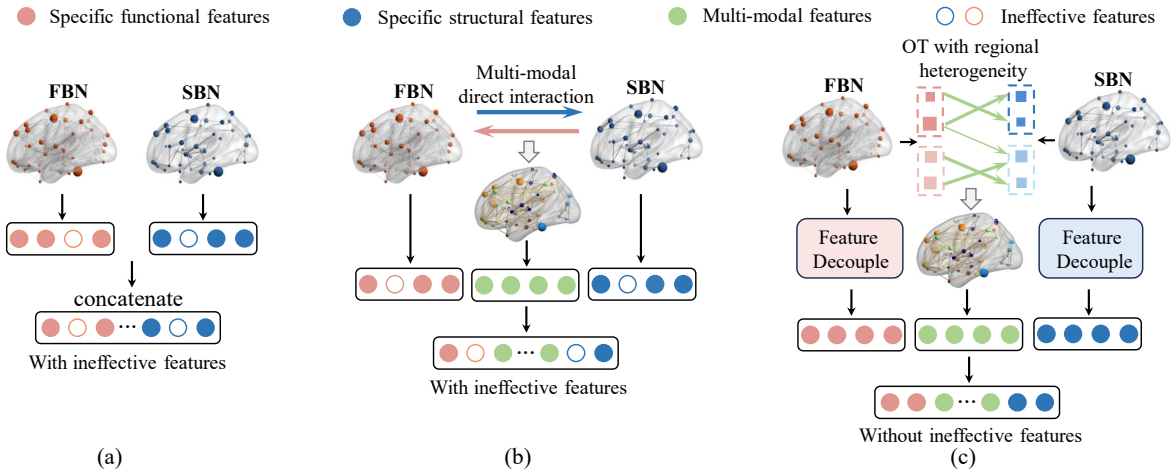


Fig. 1. Comparison of different multi-modal brain network fusion methods. (a) Simple fusion strategy. (b) Interactive fusion strategy. (c) The proposed multi-modal fusion strategy with regional heterogeneity constraints.

The anatomical structure of the brain can partially limit functional activity, and functional activity also shapes structural connectivity through neuromodulation and plasticity processes [11]. Although there is a certain correlation between structural connection strength and functional activation, the brain still possesses a wide range of indirect connections to support more complex cognitive functions. For instance, homologous visual cortices with similar visual field mappings exhibit functional connectivity even in the absence of direct structural connectivity [12]. Moreover, neuroscience studies have confirmed that the functional-structural coupling of some brain regions is more closely correlated than other regions. For example, the visual and auditory cortex show strong coupling, while parietal lobe and occipital lobe show weak coupling [13]. This regional heterogeneity reflects the brain's integration and segregation at both functional and structural levels. The brain network fusion strategy based on regional heterogeneity aligns closely with the brain's biological characteristics, facilitating a better representation of the intrinsic functional-structural interactions within the brain. Additionally, FBN and SBN also encode information unrelated to brain diseases, such as age and gender. Therefore, effectively decoupling the ineffective information is not only beneficial to improve the diagnostic accuracy, but also helps to further mine the regional heterogeneity multi-modal features related to diseases.

To address the aforementioned problems, we propose a regional heterogeneity multi-modal brain network fusion model based on optimal transport, and design a feature decoupling module to eliminate ineffective features unrelated to diseases from multi-modal brain networks. The brain is a modular system composed of multiple subnetworks, characterized by tight coupling within subnetworks and loose coupling between them [14]. Because brain regions with structural connections tend to have stronger functional synchronization, regions within the same subnetwork also tend to exhibit tighter functional-structural coupling [15], [16]. Given that the pre-defined subnetwork closely aligns with the brain structure and functional organization [17], it is beneficial to investigate the impact of brain diseases on various cognitive functions. Therefore, we first divide FBN and SBN into multiple subnetworks according to predefined clustering [18], and reorder brain regions so that brain regions belonging to the same subnetwork are adjacent, yielding regional FBN and SBN. Next, we introduce penalty weights to reduce the communication cost between cross-modal brain regions within the subnetwork and increase the cost between the subnetworks, thereby reflecting the regional heterogeneity between FBN and SBN. Then, we adopt optimal transport to simulate the transfer process of brain region hubness from FBN to SBN under the regional heterogeneous constraint. This process not only captures the intricate mapping between brain function and structure, but also effectively fuses the regional heterogeneity features between them. In addition, we design a feature decoupling module. It not only enhances the discriminability of specific and multi-modal information, but also helps the model to learn features most relevant to disease diagnosis, thereby effectively improving diagnostic performance. Finally, we fuse the effective multi-modal features and specific features for classification. The proposed multi-modal fusion strategy with regional heterogeneity constraints is shown in Fig. 1(c). In summary, the proposed method offers the following contributions:

(1) In this paper, we develop a multi-modal brain network analysis framework with regional heterogeneity constraints, which not only captures the complex association between FBN and SBN, but also alleviates the disease-unrelated information in brain networks.

(2) We design the cross-modal communication cost matrix that can reflect the regional heterogeneity, and simulate the transmission of hubness from FBN to SBN under the regional heterogeneity constraint through optimal transport. This process can effectively fuse the complex interactive information between FBN and SBN.

(3) To alleviate disease-irrelevant information in brain networks, we introduce a feature decoupling module that retains only effective multi-modal features, and apply difference constraints to enhance the discrimination of these feature representations.

(4) Extensive experiments on epilepsy and ADNI datasets demonstrate that the proposed method outperforms existing brain network analysis methods and provides multi-modal

biomarkers for brain disease diagnosis.

The remainder of the paper is organized as follows: we describe the related work in Section II. In Section III, we present the proposed multi-modal brain network analysis method in detail. Next, we introduce the two datasets used and show extensive experimental results in Section IV. After that, we discuss the experimental performance in Section V. Finally, we summarize this work and draw a conclusion.

II. Related Work

A. Optimal Transport-Based Brain Network Analysis

Optimal transport is a theoretical framework for studying probabilistic transformations, aiming to convert the original distribution to the target distribution in the most economical way. Since brain networks can be represented as probability distributions through graph theory or network embedding, optimal transport has been widely used in brain network analysis. For example, Zhu et al. utilized optimal transport to construct a graph hubness propagation model between adjacent dynamic FBNs, effectively capturing the brain's spatio-temporal evolution characteristics [6]. Ma et al. proposed an optimal transport-based ordered pattern tree kernel to measure the similarity between paired brain networks, and achieved good performance in disease classification and regression tasks [19]. Dadashkarimi et al. employed optimal transport to estimate the relationship between source and target atlas, which leads to a more accurate estimate of brain connectivities on the target atlas [20]. Numerous studies have shown that optimal transport is a powerful tool for analyzing brain networks.

However, these optimal transport-based methods focus on unimodal brain network analysis, ignoring its potential to measure information communication between multi-modal brain networks. Additionally, the functional-structural coupling between brain regions located within the same subnetwork is stronger than that between other regions. Therefore, by designing a reasonable cost matrix for brain region communication, we can capture the complex regional heterogeneous association between FBN and SBN through optimal transport.

B. Multi-Modal Features Decoupling

In brain disease diagnosis, multi-modal feature fusion refers to integrating complementary information from different brain networks to enhance feature representation. In fact, brain networks of different modalities contain both modality-specific and common information. Effectively decoupling different types of features can provide a more comprehensive view of brain disease diagnosis. In order to capture richer feature information, several multi-modal feature decoupling methods have been proposed. For example, Jiang et al. firstly constructed a specific modal graph for each data, and then established a decoupled representation model to learn both common and private information across modalities [21]. Ye et al. used subgraph convolution to extract the regional features of FBN and SBN, then applied a fusion bottleneck-based transformer to simulate the indirect interaction between them, and finally integrated these three features for brain disease diagnosis [22]. Xu et al. designed a multi-channel GCN to extract the specific and common features of FBN and SBN respectively, and introduced contrastive learning to enhance the discrimination of different features [23].

However, these methods only consider decoupling the specific features and common features from the multi-modal brain network, ignoring the disease-unrelated information encoded in the brain. Not only do brain diseases affect neural activity and brain topology, but factors such as age and gender can also alter connectivity properties [24]. Separating the ineffective information from the brain network is beneficial to provide more accurate features for brain disease diagnosis, thereby improving the diagnostic performance.

III. Proposed Method

In this section, we design a novel multi-modal brain network fusion method, which effectively captures the regional heterogeneity characteristics between FBN and SBN and suppresses the disease-irrelevant information in brain networks.The schematic diagram of the proposed method is shown in Fig. 2.

A. Regional Functional and Structural Brain Network Construction

We assume that the fMRI time series of each subject is represented by $X \in R^{N \times M}$, where N denotes the number of brain regions and M is the number of recorded time points. In order to measure the functional correlations between brain regions, we employ Pearson's correlation coefficient to construct the FBN A_f. Considering the sparsity of the brain network, we set all the elements smaller than threshold α in the FBN to zero. In addition, we denote the SBN extracted from the DTI modality as A_s. The brain is a highly regionalized system consisting of multiple subnetworks, and regional features are also shown as subnetwork features [22]. In order to obtain the regional FBN and SBN, we divide each brain network into multiple subnetworks according to the predefined clustering [18], including default mode network (DMN), frontoparietal network (FPN), salience network (SAN), attention network (ATN), somatosensory and motor network (SMN), visual network (VN), and auditory network (AN). Then, we reorder the brain regions so that those belonging to the same subnetwork are located in adjacent positions. Subsequently, regional FBN and SBN are formed in an approximately block-diagonal manner.

B. Regional Heterogeneous Fusion with Optimal Transport

To capture the complex coupling between functional and structural networks, we develop a cross-modal hubness propagation model with regional heterogeneity constraints via optimal transport. Specifically, we first introduce penalty weights to reduce the communication cost of cross-modal brain region interactions within subnetworks and increase the cost between different subnetworks, thereby reflecting the regional heterogeneity between FBN and SBN. Then, we utilize optimal transport to simulate the hubness propagation process from the FBN to the SBN under the regional heterogeneity constraint.

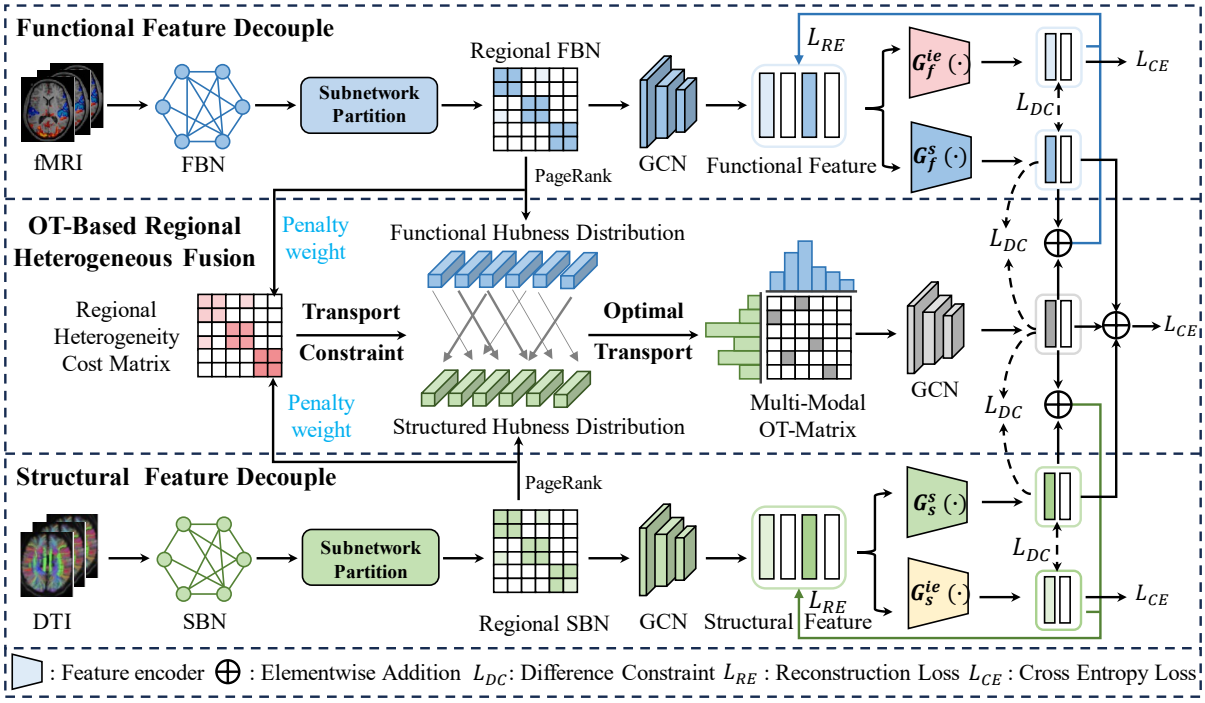


Fig. 2. The framework of the proposed novel multi-modal brain network analysis method. In the regional FBN and SBN, diagonal blocks represent different subnetworks. PageRank is used to compute the hubness distributions of the regional FBN and SBN, which serve as the initial distribution and target distribution for the optimal transport process, respectively. The transport constraint refers to the cost matrix, which ensures that the optimal transport transforms the initial distribution into the target distribution with minimal cost.

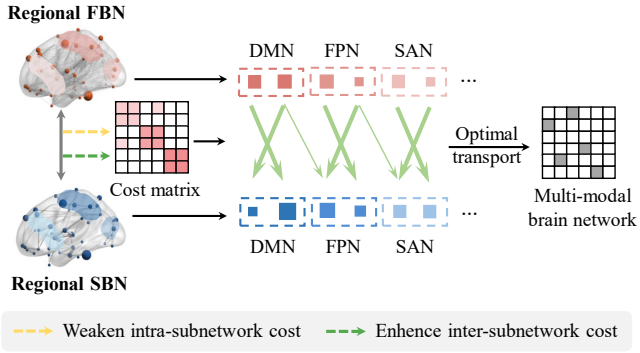


Fig. 3. Schematic diagram of regional heterogeneous multi-modal brain network analysis based on optimal transport. The cost matrix represents the cost required to transport hubness from functional brain regions to structural brain regions. The solid green arrows indicate the hubness transport process between functional and structural brain regions. DMN, FPN, and SAN represent different subnetworks.

This process can effectively capture the intricate interactions between functional and structural brain regions. The multi-modal fusion process based on optimal transport is shown in Fig. 3.

Measure of Node Hubness: Since PageRank [25] can efficiently measure the node importance according to the brain topology, we use it to calculate the hubness of each brain region in regional FBN and regional SBN. The hubness h^u of the brain region u depends on the number and hubness of its neighbors:

$$h^u = \sum_{v \in N(u)} \frac{h^v}{D_v} \tag{1}$$

where $N(u)$ denotes the set of neighbors of the brain region u, and D_v is the degree of brain region v. Therefore, we can obtain that the hubness distribution of regional FBN is $h_f = (h_f^1, h_f^2, \cdots, h_f^N) \in R^{1\times N}$, and the hubness distribution of regional SBN is $h_s = (h_s^1, h_s^2, \cdots, h_s^N) \in R^{1\times N}$.

Construction of Regional Heterogeneity Cost Matrix: In order to characterize the communication efficiency between functional and structural brain regions, we first use cosine similarity to measure the correlation between cross-modal brain regions:

$$B_{ij} = \frac{< A_f^i, A_s^j >}{\|A_f^i\| \|A_s^j\|} \tag{2}$$

where B_{ij} represents the correlation between the i^{th} functional brain region and the j^{th} structural brain region, $< A_f^i, A_s^j >$ is the inner product of A_f^i and A_s^j, and $\|\cdot\|$ denotes the Frobenius norm. Neuroscience studies have shown that there is regional heterogeneity in the function-structure coupling, characterized by intensive information exchange between cross-modal brain regions within the same subnetwork, and sparse information transfer between different subnetworks [15], [16]. In order to align with the brain's intrinsic topology, we introduce penalty weights to enhance connections within subnetworks and weaken those between them:

$$C_{ij} = \begin{cases} W_{intra} \times B_{ij} & i, j \in \textit{same subnetwork} \\ W_{inter} \times B_{ij} & \textit{otherwise} \end{cases} \tag{3}$$

where W_{intra} and W_{inter} are penalty weights, which are used to strengthen and weaken correlations, respectively. A higher correlation between functional and structural brain regions suggests a lower communication cost across modal brain regions. To this end, we introduce a penalty factor related

to the strength of cross-modal coupling, and the resulting communication cost matrix P for the regional heterogeneity constraint can be formulated as follows:

$$P_{ij} = e^{(-\frac{C_{ij}^2}{\sigma})} \tag{4}$$

where σ is used to adjust the weight decay speed for the cross-modal connectivity strength.

Multi-Modal Fusion with Optimal Transport: We use the cost matrix constrained by heterogeneity as the transport cost and apply optimal transport to simulate functional-structural interactions at minimal cost. This process not only facilitates the integration of multi-modal information with regional heterogeneity but also aligns with the brain's complex economic characteristics. Specifically, we take the functional hubness distribution as the source domain, the structural hubness distribution as the target domain, and the communication cost P as the transport cost to establish the cross-modal hubness propagation model with regional heterogeneity constraints:

$$\begin{aligned} &\min_{M \in R^{N \times N}} < P, M >_F, \\ &s.t. M\mathbf{1}_N = h_f, M^T\mathbf{1}_N = h_s, \\ &M\mathbf{1}_N = \Sigma_j M_{ij}, M^T\mathbf{1}_N = \Sigma_i M_{ij} \end{aligned} \tag{5}$$

where $< \cdot, \cdot >_F$ is the Frobenius inner product of matrices, $< P, M >_F = \Sigma_i \Sigma_j P_{ij} M_{ij}$, M_{ij} denotes the hubness transport from h_f^i to h_s^j, and $\mathbf{1}_N$ represents the N-dimensional all-ones vector. Optimal transport aims to transfer the hubness of functional brain regions to multiple structural brain regions at the lowest cost. Moreover, the total hubness transmitted from each functional brain region to structural brain regions equals its initial hubness, while the total hubness received by each structural brain region equals its target hubness. This low-cost many-to-many interaction is highly consistent with the brain's high-order property and economic principle [26], [27]. Therefore, the proposed method is beneficial for capturing the intrinsic interactions between functional and structural brain networks, enabling more effective integration of multi-modal brain network information. In this paper, we use Sinkhorn algorithm [28] to solve the optimal transport problem. Therefore, the obtained optimal transport plan M can effectively fuse the regional heterogeneity characteristics between FBN and SBN.

In order to extract the rich spatial structure information in multi-modal brain networks with regional heterogeneity, we take the SBN A_s as the adjacency matrix and M as the node feature, and adopt GCN [29] to learn the complex interactions representations between brain regions:

$$Z^l = Relu(D_s^{-1/2} A_s D_s^{-1/2} Z^{l-1} W^l) \tag{6}$$

where Z^l ($Z^0 = M$) is the topological feature obtained after l-layer graph convolution, *Relu* is the nonlinear activation function [30], [31], D_s represents the degree matrix of A_s, and W^l denotes the learnable weight matrix in the l-layer graph convolution.

C. Feature Decoupling Module

The brain network also encodes identity information [24], such as age and gender, which will interfere with the brain disease diagnosis. To address this issue, we design a feature decoupling module aiming to separate the specific information related to the disease from the ineffective information unrelated to the disease in both the FBN and SBN. Specifically, we first employ GCN to extract the nonlinear structural features of regional FBN and SBN, respectively:

$$Z_i^l = Relu(D_i^{-1/2} A_i D_i^{-1/2} Z_i^{l-1} W_i^l) \tag{7}$$

where $i \in \{f, s\}$, $Z_i^l (Z_f^0 = X, Z_s^0 = A_s)$ represents the feature embedding obtained after l-layer graph convolution, D_i is the degree matrix of A_i, and W_i^l denotes the learnable weight matrix in the l-layer graph convolution.

To decouple the specific and ineffective features in FBN and SBN, we design two encoders (e.g., GCN) G_i^s and G_i^{ie} ($i \in \{f, s\}$) for each brain network to obtain the latent representation of the two types of features:

$$H_i^s = G_i^s(Z_i), H_i^{ie} = G_i^{ie}(Z_i) \tag{8}$$

Then, the difference constraint L_{DC1} is used to remove the correlation between specific features and ineffective features:

$$L_{DC1} = CS\big(H_f^s, H_f^{ie}\big) + CS\big(H_s^s, H_s^{ie}\big) \tag{9}$$

where $CS(\cdot, \cdot)$ denotes the cosine similarity. In addition, to further identify ineffective features, we feed H_i^{ie} into a multi-layer perceptron (MLP) to predict the class of each subject and maximize the cross-entropy loss $CE(\cdot, \cdot)$ between the true label y and the predicted label:

$$L_{CE1} = CE\left(y, MLP\left(H_f^{ie}\right)\right) + CE\left(y, MLP\left(H_s^{ie}\right)\right) \tag{10}$$

In order to improve the discrimination between specific and multi-modal features, we also introduce difference constraints between them:

$$L_{DC2} = CS\Big(H_f^s, Z\Big) + CS(H_s^s, Z) \tag{11}$$

Next, we fuse ineffective, specific and multi-modal features to generate reconstructed features, and introduce the MSE loss [32] to constrain the similarity between the reconstructed feature representation and the original feature representation:

$$L_{RE} = \| H_f^{ie} \oplus H_f^s \oplus Z - Z_f \|_2^2 + \| H_s^{ie} \oplus H_s^s \oplus Z - Z_s \|_2^2 \tag{12}$$

where $\|\|_2^2$ denotes the squared L2 norm, and $\oplus$ is the element-wise addition. We introduce difference constraints and reconstruction loss to ensure that the decoupled disease-relevant modality-specific features, disease-irrelevant modality-specific features, and multi-modal features are distinct yet complementary, effectively reducing disease-irrelevant information in both modality-specific and multi-modal features.

Finally, we sum the function-specific features, structure-specific features, and multi-modal features into the MLP for classification, and minimize the cross-entropy loss between the true label y and the predicted label:

$$L_{CE2} = CE\left(y, MLP(H_f^s \oplus H_s^s \oplus Z)\right) \tag{13}$$

D. Model Optimization

In the training process, the loss function of the proposed model consists of three components: cross-entropy classification loss, difference constraint loss and reconstruction loss. where cross-entropy loss $L_{CE} = L_{CE2} - L_{CE1}$, and difference constraint loss $L_{DC} = L_{DC1} + L_{DC2}$. Therefore, the overall loss function can be expressed as follows:

$$L = L_{CE} + \lambda_1 L_{DC} + \lambda_2 L_{RE} \tag{14}$$

where λ_1 and λ_2 are hyperparameters that determine the relative importance of the three types of losses.

IV. Experiment

A. Materials

Epilepsy and cognitive impairment are among the most damaging neurological disorders, disrupting brain structure and leading to neurological abnormalities. Therefore, studying their pathological mechanisms and diagnostic methods is of great significance for early intervention and treatment. Next, we will present the details of the two datasets used.

1) Data Acquisition: We conduct extensive experiments on fMRI and DTI data from the collected epilepsy dataset and the publicly available ADNI dataset (https://adni.loni.usc.edu/). **Epilepsy:** This dataset is collected by Jinling Hospital, Nanjing University School of Medicine, and contains a total of 306 subjects, including 114 healthy controls (HC), 103 frontal lobe epilepsy patients (FLE), and 89 temporal lobe epilepsy patients (TLE). The data acquisition equipment is Siemens Trio 3T scanner. The scanning parameters of fMRI are: repetition time = 2000ms, echo time = 30 ms, flip angle = 90°, and voxel size $= 3.75 \times 3.75 \times 3.75 mm^3$. The scanning parameters of DTI are: repetition time = 6100 ms, echo time = 93 ms, flip angle = 90°, and voxel size $= 0.94 \times 0.94 \times 0.94 mm^3$, respectively. **ADNI:** We download a total of 227 samples from ADNI2 and ADNI3, including 82 healthy controls (HCs), 76 patients with significant memory concern (SMC), and 69 patients with early mild cognitive impairment (EMCI). The HC group consists of 40 males and 42 females, the SMC group includes 50 males and 26 females, and the EMCI group comprises 37 males and 32 females.

2) Data Preprocessing: We use the same way to process the epilepsy dataset and the ADNI dataset. Specifically, all fMRI data are processed by SPM8 in DPABI toolbox. We first split the serialized data into multiple pieces and adjust them according to the EPI template, thus achieving correction and alignment of raw images. Next, detrending is used to reduce the effects of head motion and the interference of CSF and white matter. After that, we employ the Automated Anatomical Labeling (AAL) atlas [33] to divide the fMRI of the epilepsy dataset into 90 brain regions with 240 time points, and the fMRI of the ADNI dataset into 90 brain regions with 197 time points. For DTI data, we first employ FSL to correct DTI distortions, then generate fiber images using TrackVis, and designate anatomical regions based on the subject's T1 images using the AAL atlas. Finally, the number of fibers obtained is used as the strength of the structural connections.

B. Comparison Methods

To measure the effectiveness of the proposed method, we compare it with 13 state-of-the-art brain network analysis methods. Among them, DTI-based methods are: convolutional neural networks framework for brain networks (BNetCNN) [34], graph convolutional neural networks (GCNN) [35] and multi-task graph structure learning framework (MNC-Net) [36]. FMRI-based methods are: graph attention networks (GAT) [37], brain network transformer (BNT) [38], functional brain network generation (FBNetGen) [39], spatio-temporal fMRI (ST-fMRI) [40] and learnable subdivision graph neural network (LSGNN) [41]. DTI&fMRI-based methods are: weighted summation (W-Sum) [42], feature concatenation (F-Concat) [43], multi-kernel Learning (MKL) [44], multi-modal graph disentangled representation (MGDR) [21], adaptive multi-channel graph convolutional networks (AM-GCN) [45], cross-modal graph neural network (Cross-GNN) [10], transformer-based self-supervised graph reconstruction framework (TSGR) [46] and regional heterogeneous multi-modal brain networks fusion strategy (RH-Brains) [22].

(1) BNetCNN: This method designs edge-to-edge, edge-to-node, and node-to-graph convolutional filters to capture the local topological properties in the SBN.

(2) GCNN: This method designs a multi-class graph convolutional classifier to extract the non-euclidean structure in SBNs.

(3) MNC-Net: This method designs a graph learning strategy based on node clustering to reduce the influence of irrelevant noise on brain disease diagnosis.

(4) GAT: This method can learn different weights for different nodes in the neighborhood, enabling the adaptive aggregation of structural features in brain graphs.

(5) BNT: This method uses FBN as node features and location information, and learns the connection strength between brain regions based on transformer.

(6) FBNetGen: This method can transform the raw fMRI time signals into task-oriented functional brain networks, and it is interpretable.

(7) ST-fMRI: This method designs a novel spatio-temporal graph convolutional framework to explicitly model long-range dynamic interactions in DFBNs.

(8) LSGNN: This method first encodes the brain network into multiple feature subnetworks corresponding to the functional configuration, and extracts the spatial structure features of each subnetwork separately.

(9) W-Sum: This method first performs a weighted sum of FBN and SBN, and then uses a multilayer perceptron for disease prediction.

(10) F-Concat: This method first concatenates FBN and SBN, and then uses multi-layer perceptron for disease prediction.

(11) MKL: This method maps SBN and FBN to a high-dimensional space by kernel functions and applies linear fusion.

(12) MGDR: This method constructs a specific modal graph for each data, and then establishes a decoupled representation model to learn both common and private information across modalities.

TABLE I
CLASSIFICATION PERFORMANCE (MEAN±STD) OF THE PROPOSED METHOD AND ALL COMPARISON METHODS ON THE EPILEPSY DATASET (%).

Modality	Method	HC vs. Patient			HC vs. FLE			HC vs. TLE			TLE vs. FLE		
		ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
DTI	BNetCNN	71.27±06.71	79.24±06.15	62.15±09.63	71.97±07.23	65.80±19.59	68.10±10.44	74.19±05.60	57.55±26.05	66.98±09.12	71.84±04.91	72.16±10.05	68.12±09.65
	GCNN	71.25±04.24	78.86±05.78	61.98±06.21	71.43±08.65	54.30±31.89	63.93±15.29	73.24±07.86	67.48±17.80	68.94±10.76	72.47±08.98	67.44±25.99	68.51±12.26
	MNC-Net	73.00±06.40	79.34±08.41	66.76±10.26	73.50±03.91	61.25±19.38	63.90±09.73	73.50±08.38	64.47±13.04	65.46±08.84	71.88±11.61	70.95±20.82	60.39±17.69
fMRI	GAT	73.85±06.14	79.94±04.83	70.83±09.31	75.19±08.80	71.02±16.85	71.58±18.02	73.79±09.41	57.53±12.54	71.64±09.52	69.76±10.26	74.53±12.15	66.65±09.38
	BNT	69.29±07.53	74.50±08.48	64.44±09.60	72.00±06.78	67.11±11.40	52.00±16.39	75.00±07.42	66.03±13.58	68.14±10.92	69.50±06.69	63.04±27.32	52.23±14.78
	FBNetGen	74.80±06.99	80.18±05.78	70.52±11.09	72.00±07.48	71.87±07.41	64.13±15.23	70.63±08.86	68.45±14.09	59.72±11.89	68.75±07.40	69.16±11.00	58.20±12.97
	ST-fMRI	73.19±06.93	81.08±05.77	63.35±15.11	75.22±08.52	71.40±11.12	70.41±12.90	76.76±04.82	66.57±15.52	75.49±09.67	70.34±04.46	74.59±05.17	70.94±09.98
	LSGNN	71.57±05.27	80.27±04.82	62.04±12.24	70.48±08.13	63.74±22.84	68.86±11.28	68.38±12.45	58.22±18.49	65.46±13.82	65.58±05.53	67.09±09.48	60.02±11.43
DTI& fMRI	W-Sum	73.56±06.97	78.87±07.69	72.34±07.69	74.72±07.50	69.77±11.07	71.89±11.51	76.26±08.34	56.77±28.55	70.66±12.05	71.32±07.12	68.69±14.55	67.69±13.13
	F-Concat	72.85±04.51	78.32±04.89	70.95±05.79	73.33±07.30	64.75±19.33	71.81±06.63	72.76±06.43	62.17±21.98	71.40±09.44	68.74±10.55	68.78±16.83	63.35±14.22
	MKL	68.91±07.56	74.40±09.38	72.87±09.27	65.76±09.09	64.21±09.18	71.27±12.35	67.29±11.99	62.70±13.53	65.52±13.08	54.11±09.29	58.41±09.52	58.80±09.87
	GMDR	74.88±09.97	79.70±10.33	68.63±10.33	75.17±08.13	67.28±15.84	73.44±12.11	74.74±11.06	68.24±16.81	72.14±13.43	72.42±05.05	72.03±11.96	67.55±09.49
	AM-GCN	76.50±06.34	80.63±04.45	65.65±12.54	75.09±09.25	63.38±21.77	67.67±15.59	75.86±06.46	67.48±12.48	69.09±11.55	71.45±09.66	70.02±13.07	65.35±15.39
	Cross-GNN	75.13±04.41	82.30±03.11	72.81±07.35	76.52±05.47	70.67±09.04	**76.29±08.91**	75.24±03.90	71.97±04.78	78.97±06.94	71.34±06.35	73.96±04.42	66.80±12.14
	TSGR	74.49±05.84	80.05±06.25	70.80±10.34	74.72±06.65	71.20±12.53	71.40±11.99	75.26±06.00	68.40±15.53	72.98±11.06	72.48±05.43	**76.04±05.65**	71.01±08.94
	RH-BrainFS	75.55±06.90	80.78±07.07	70.71±11.46	73.77±06.96	66.81±17.87	67.38±13.51	76.50±07.43	66.05±14.71	69.72±09.81	72.50±05.73	65.45±25.02	65.25±16.69
	Ours	**79.04±06.69**	**82.80±07.64**	**76.62±07.45**	**77.90±08.33**	**76.93±06.84**	71.97±10.98	**82.71±06.85**	**76.69±14.92**	**84.47±10.39**	**74.50±05.34**	73.95±13.69	**73.07±09.44**

(13) AM-GCN: This method introduces a multi-channel GCN to extract the specific information and shared information of SBN and FBN, respectively, and introduces similarity constraints and difference constraints to enhance the discrimination of different feature representations.

(14) Cross-GNN: This method first obtains the explicit complementary features of FBN and SBN through the inter-modal correspondence matrix, and then uses the self-knowledge distillation mechanism to learn the implicit correlation between them.

(15) TSGR: This method uses GCN to fuse topological features in FBN and SBN, and reconstructs the brain map through a self-supervised model to identify key information.

(16) RH-BrainFS: This method introduces a subgraph convolution to extract the regional features of FBN and SBN, and further uses a Transformer-based fusion bottleneck to fuse the indirect interaction between them.

C. Experimental Setup

The proposed model is trained on an NVIDIA GeForce RTX 3080 GPU. During training, the number of epochs is 300 and batch size is 20. We use the Adam [47] with a learning rate of 1e-3 to optimize the proposed method. In the experiments, we use grid search to determine the different hyperparameters. Specifically, the brain network threshold α ranges from 0.1 to 0.9 with a step size of 0.05. The intra-subnetwork penalty weight W_{intra} is $\{1.2, 1.4, 1.5, 1.6, 1.8\}$, and the inter-subnetwork penalty weight W_{inter} is varied within the range of $\{0.2, 0.4, 0.5, 0.6, 0.8\}$. The experimental results show that the diagnostic performance is optimal when W_{inter}/W_{intra} is between 3 and 4. This may be because the brain network constructed under such conditions not only enhances efficient interactions between nodes within subnetworks, but also maintains appropriate connectivity between subnetworks to support cross-subnetwork collaboration. λ_1 and λ_2 in Eq. (14) vary within the range of $\{$1e-3, 1e-2, 1e-1, 1, 10$\}$. In the network framework, all feature encoders are implemented as 2-layer GCNs, and the number of neurons in the three fully connected layers is 256, 64 and 2, respectively.

In the experiments, we employ ten-fold cross-validation to evaluate the performance of all methods. Specifically, we divide the whole dataset into 10 approximately equal but mutually exclusive subsets, with 9 subsets used for training and 1 subset for testing. The experiment is repeated ten times, and we record the mean and standard deviation of all test results. To prevent overfitting, we introduce the early stopping method. Futhermore, In order to ensure the stability and effectiveness of model training, we set the dropout to 0.5 and introduce a gradient clipping strategy to promote model convergence. When the training loss does not decrease for 30 consecutive epochs, the model parameters at this time are saved for testing. For the epilepsy dataset, we design four classification tasks: HC vs. Patients (FLE&TLE), HC vs. FLE, HC vs. TLE and FLE vs. TLE. For the ADNI dataset, we also set up four classification tasks: HC vs. Patient (SMC&EMCI), HC vs. SMC, HC vs. EMCI and SMC vs. EMCI. All classification tasks are evaluated by accuracy (ACC), f1-score (F1), and the area under the receiver operation characteristic curve (AUC).

D. Classification Performance

Table I and Tabel II show the diagnostic performance of different methods on epilepsy and ADNI datasets, respectively. From these tables, we can find that the proposed method significantly outperforms other comparison methods. Specifically, for the epilepsy dataset, the proposed method

TABLE II
CLASSIFICATION PERFORMANCE (MEAN±STD) OF THE PROPOSED METHOD AND ALL COMPARISON METHODS ON THE ADNI DATASET (%).

Modality	Method	HC vs. Patient			HC vs. SMC			HC vs. EMCI			SMC vs. EMCI		
		ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
DTI	BNetCNN	80.64±14.89	84.06±12.63	85.55±11.12	75.52±16.25	78.36±10.83	79.53±11.01	81.26±08.31	76.46±13.87	83.52±10.58	77.69±07.26	78.41±08.06	72.56±09.40
	GCNN	75.65±07.40	79.92±07.19	76.38±16.71	81.58±07.29	77.06±12.50	82.47±10.80	77.33±08.54	67.03±22.98	71.02±14.89	75.08±12.04	80.25±08.52	78.87±11.65
	MNC-Net	77.50±12.09	85.13±08.11	86.48±07.69	84.38±11.07	83.04±11.04	88.89±08.45	85.83±06.51	81.56±12.55	85.17±09.35	81.67±07.26	75.14±17.92	78.27±13.55
fMRI	GAT	81.86±08.33	68.30±13.46	79.80±11.51	84.71±09.71	81.80±11.34	86.12±09.60	80.67±14.13	73.44±18.58	78.06±14.34	80.62±08.47	80.84±08.53	80.88±09.66
	BNT	82.00±06.40	84.63±07.04	66.60±23.97	86.67±07.64	85.42±09.41	76.02±20.17	79.17±10.70	71.30±26.51	78.76±11.51	78.33±06.67	69.84±27.99	64.94±15.28
	FBNetGen	80.63±07.63	82.55±09.60	62.02±15.32	86.25±06.12	84.20±07.93	78.90±11.44	78.33±13.54	76.29±15.86	65.67±17.93	76.67±08.98	72.29±16.23	58.30±27.04
	ST-fMRI	81.34±07.84	85.26±07.29	77.41±09.35	85.50±06.81	82.79±09.04	86.14±10.02	84.00±09.52	80.44±11.74	80.70±14.68	79.24±06.80	77.14±15.96	76.93±14.84
	LSGNN	72.53±04.08	80.30±05.31	66.95±09.28	70.38±07.60	66.10±15.70	62.45±17.75	69.33±10.83	45.23±33.29	68.06±11.06	71.33±12.67	54.87±31.20	64.08±19.07
DTI& fMRI	W-Sum	73.50±04.50	73.70±08.06	85.04±05.70	78.33±15.46	73.06 ±20.00	70.80±18.75	85.83±11.21	79.33±20.14	82.66±18.29	79.17±10.70	77.40±12.28	74.35±20.30
	F-Concat	80.45±10.63	80.29±13.20	94.62±06.58	85.00±12.80	85.81±09.06	81.74±16.69	81.67±07.26	71.35±14.17	76.27±15.94	76.67±13.33	75.21±15.06	76.77±13.73
	MKL	68.38±08.94	66.23±11.38	80.14±12.04	83.96±10.51	83.35±12.84	86.80±11.72	77.88±07.84	59.45±23.08	73.51±19.37	71.54±00.92	65.82±10.82	72.80±14.27
	MGDR	77.24±07.56	75.42±14.71	81.43± 05.91	80.48±09.69	74.76± 14.97	82.96±13.58	76.76±09.58	58.85±23.91	74.65± 16.51	78.46±12.31	71.22±20.88	76.51±19.10
	AM-GCN	88.19±06.67	88.93±07.38	90.13±07.45	86.19±11.78	79.03±18.60	85.62±15.21	87.69±08.57	84.07±14.19	**91.50±08.81**	84.89±09.62	84.16±11.55	85.17±09.57
	Cross-GNN	86.14±06.69	88.42±06.28	90.56±07.04	88.33±11.87	84.17±17.79	89.26±10.16	86.43±06.60	77.31±19.98	82.48±13.25	83.08±04.62	80.13±14.05	77.50±15.39
	TSGR	86.64±05.08	89.58±04.26	92.48±05.74	87.50±06.72	84.97±09.10	91.49±06.99	83.33±12.91	76.22±19.75	84.66±15.59	83.33±08.33	84.30±10.90	85.88±09.71
	RH-BrainFS	85.63±07.93	87.34±07.84	84.45±06.77	85.83±09.17	83.43±10.07	86.49±11.14	84.17±05.83	82.84±04.72	73.77±15.32	82.50±06.92	78.51±09.38	86.08±09.34
	Ours	**91.14±06.07**	**92.41±05.08**	**94.87±05.62**	**90.19±06.43**	**87.71±09.11**	**91.61±08.69**	**89.23±06.15**	**86.31±09.16**	88.99±07.00	**88.02±08.36**	**87.22±09.88**	**87.39±11.88**

TABLE III
ABLATION EXPERIMENTS UNDER DIFFERENT CLASSIFICATION TASKS ON EPILEPSY AND ADNI DATASETS (%).

Dataset	Method	HC vs. Patient			HC vs. FLE			HC vs. TLE			FLE vs. TLE		
		ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
Epilepsy	baseline	66.73	74.30	55.05	68.18	63.02	65.07	74.19	51.87	63.69	68.13	60.89	58.14
	w/o regional heterogeneity	75.78	79.88	70.72	73.27	67.80	66.25	78.83	72.86	74.73	71.87	72.62	64.70
	w/o optimal transport	76.25	82.48	72.76	75.63	66.30	61.93	80.63	72.34	74.19	71.25	70.77	65.76
	w/o feature decouple	73.55	79.23	65.14	73.71	64.91	63.77	78.50	75.56	76.98	73.75	72.51	70.74
	ours	**79.04**	**82.80**	**76.62**	**77.90**	**76.93**	**71.97**	**82.71**	**76.69**	**84.47**	**74.50**	**73.95**	**73.07**
		HC vs. Patient			HC vs. SMC			HC vs. EMCI			SMC vs. EMCI		
ADNI	baseline	70.88	67.30	80.78	80.48	74.76	82.22	75.99	56.90	75.16	76.92	74.08	80.73
	w/o regional heterogeneity	89.21	90.95	91.62	86.81	83.93	**92.58**	88.79	84.39	88.10	84.95	83.97	86.11
	w/o optimal transport	88.50	89.85	89.82	86.19	84.78	86.33	87.50	83.30	90.44	86.67	84.83	89.78
	w/o feature decouple	88.05	89.11	87.97	86.67	78.94	82.24	87.50	80.83	**95.00**	84.16	82.50	86.90
	ours	**91.14**	**92.41**	**94.87**	**90.19**	**87.71**	91.61	**89.23**	**86.31**	88.99	**88.02**	**87.22**	**87.39**

w/o means without.

achieves accuracy improvements of 2.54%, 1.38%, 5.95% and 2.00% across the four classification tasks, respectively. For the ADNI dataset, the proposed method achieves accuracy improvements of 2.95%, 1.86%, 1.54%, and 3.13% across the four classification tasks, respectively. Notably, the proposed method also achieves almost optimal performance on other metrics, providing strong evidence of its effectiveness.

The reason why our method performs better is that it not only captures the regional heterogeneous interaction between FBN and SBN, but also removes the disease-irrelevant features from brain networks. Unlike single-modal methods such as BNetCNN, BNT and LSGNN, which ignore the information fusion between different brain networks and result in suboptimal diagnostic performance. Compared with multi-modal methods such as MKL, Cross-GNN, and TSGR, they overlook the fact that some brain regions have stronger function-structure coupling than others. Therefore, they fail to comprehensively explore the nonlinear associations between these networks. In contrast, our method effectively simulates the regional heterogeneity mapping between FBN and SBN using optimal transport, and introduces a feature decoupling module to remove ineffective information, thereby enhancing the discriminability of brain disease-related features.

E. Ablation Study

The proposed method consists of three important modules: regional heterogeneity, optimal transport, and feature decoupling. To evaluate the contribution of each module, we conduct ablation experiments on the epilepsy and ADNI datasets, and the experimental results are shown in Table III. The simplified models used in the experiments include:

TABLE IV
RESULTS OF INDEPENDENT ABLATION EXPERIMENTS FOR DIFFERENT MODULES (%).

Dataset	Method	HC vs. Patient			HC vs. FLE			HC vs. TLE			FLE vs. TLE		
		ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
Epilepsy	baseline	66.73	74.30	55.05	68.18	63.02	65.07	74.19	51.87	63.69	68.13	60.89	58.14
	method with regional heterogeneity	71.42	78.39	64.53	73.00	66.98	66.71	74.50	69.91	72.35	68.75	69.57	64.65
	method with optimal transport	70.71	77.55	67.58	72.00	64.06	63.68	75.50	69.62	71.50	70.88	72.53	65.90
	method with feature decouple	73.93	78.62	64.61	71.50	65.02	66.68	75.50	67.82	70.83	71.12	69.76	65.26
	ours	**79.04**	**82.80**	**76.62**	**77.90**	**76.93**	**71.97**	**82.71**	**76.69**	**84.47**	**74.50**	**73.95**	**73.07**
		HC vs. Patient			HC vs. SMC			HC vs. EMCI			SMC vs. EMCI		
ADNI	baseline	70.88	67.30	80.78	80.48	74.76	82.22	75.99	56.90	75.16	76.92	74.08	80.73
	method with regional heterogeneity	86.00	85.83	91.14	85.83	83.43	91.25	85.83	81.29	**93.48**	83.33	81.08	83.21
	method with optimal transport	86.50	88.72	90.25	85.00	82.09	90.67	86.67	78.68	93.27	80.83	75.44	80.69
	method with feature decouple	87.50	88.31	88.17	85.83	80.60	84.63	85.63	81.19	89.88	82.50	82.51	82.43
	ours	**91.14**	**92.41**	**94.87**	**90.19**	**87.71**	**91.61**	**89.23**	**86.31**	88.99	**88.02**	**87.22**	**87.39**

(1) baseline: SBN and FBN are used as adjacency matrix and node features, respectively, and GCN is used to extract multi-modal fusion information.

(2) w/o regional heterogeneity: Both W_{intra} and W_{inter} in Eq. (3) are set to 1.

(3) w/o optimal transport: The multi-modal optimal transport matrix is replaced by the function-structural correlation matrix in Eq. (3).

(4) w/o feature decouple: The difference constraint is removed.

The experimental results in Table III present that each module significantly improves the diagnostic performance compared to the baseline. Firstly, removing regional heterogeneity leads to a decline in performance. This is because the functional-structural coupling in some brain regions is tightly correlated, and regional heterogeneity constraints can more accurately describe the intrinsic functional-structural interactions. Second, the multi-modal optimal transport matrix achieves better results than the function-structural correlation matrix, as the relationship between FBN and SBN is nonlinear. The optimal transport helps capture the complex high-order mapping between them by simulating hubness transmission from FBN to SBN. In addition, the feature decoupling module effectively alleviates information unrelated to the brain disease and enhances the discrimination between different feature representations. Finally, the classification performance is optimal when all modules are used in combination. These results suggest that the proposed modules are closely related and promote each other, which improves the recognition performance.

To verify the unique contributions of each module and the interactions between them, we conduct independent ablation experiments for different modules. The experimental results are shown in Table IV. The simplified models used in the experiments include:

(1) baseline: SBN and FBN are used as adjacency matrix and node features, respectively, and GCN is used to extract multi-modal fusion information.

(2) method with regional heterogeneity: The multi-modal optimal transport matrix is replaced by the function-structure coupling matrix in Eq. (3) , and the λ_1 and λ_2 in Eq. (14) are set to 0.

(3) method with optimal transport: Both W_{intra} and W_{inter} in Eq. (3) are set to 1, and λ_1 and λ_2 in Eq. (14) are set to 0.

(4) method with feature decouple: Both W_{intra} and W_{inter} in Eq. (3) are set to 1, and the multi-modal optimal transport matrix is replaced by the function-structure coupling matrix in Eq. (3).

From Table IV, we can draw the following conclusions. First, compared with baseline, each independent module can improve the diagnostic performance, proving the effectiveness of the three modules we proposed. In addition, compared with Table III, we can observe that the diagnosis accuracy is higher when the two modules are used jointly than when only one module is used. These phenomena further present that the three modules all contribute positively to enhancing disease diagnostic performance and that they promote each other.

V. DISCUSSION

A. Parameter Sensitivity Analysis

There are two hyperparameters λ_1 and λ_2 in the objective function, which are used to control the relative importance of the classification loss, the difference constraint, and the reconstruction loss. Both λ_1 and λ_2 vary within the range of $\{$1e-3,1e-2,1e-1,1,10$\}$. In Fig. 4, we explore the impact of different parameter combinations on diagnostic performance. The experimental results show that the classification accuracy remains stable within the candidate range and does not fluctuate significantly with changes in λ_1 and λ_2. Therefore, the proposed method is robust and insensitive to hyperparameters.

B. Embedding Features Visualization

To present the feature representation abilities of different methods, we adopt T-SNE to map high-dimensional features to a low-dimensional space for visualization. As shown in Fig. 5, compared to the original data with chaotic distribution, Cross-GNN is able to cluster intra-class data but struggles to identify shared features across different groups of subjects. RH-BrainFS can separate inter-class data but cannot effectively cluster intra-class data. This is because these methods fail to accurately describe the mapping relationship between FBN and SBN, resulting in weak discrimination of feature

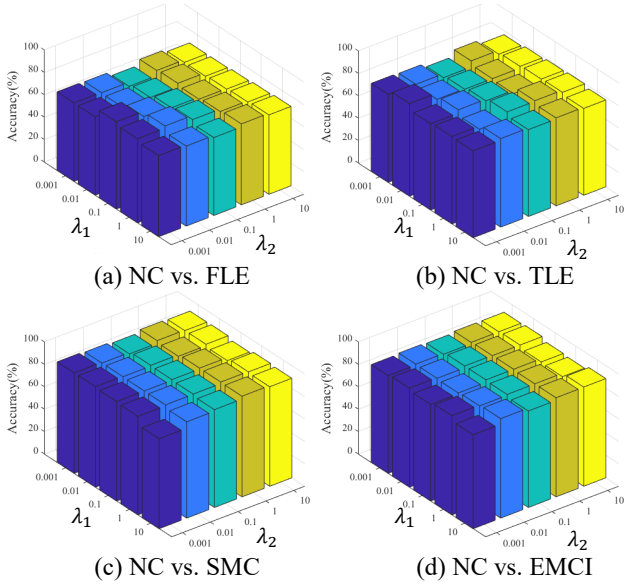


Fig. 4. The influence of λ_1 and λ_2 on different classification tasks. λ_1 and λ_2 represent hyperparameters used to control the relative importance of classification loss, discrepancy constraint loss, and reconstruction loss. The Z-axis represents the classification accuracy under different combinations of these parameters.

representation. Notably, our method captures the regional heterogeneous interactions between function and structure, and also removes the disease-unrelated information, thus effectively distinguishing the features of different categories.

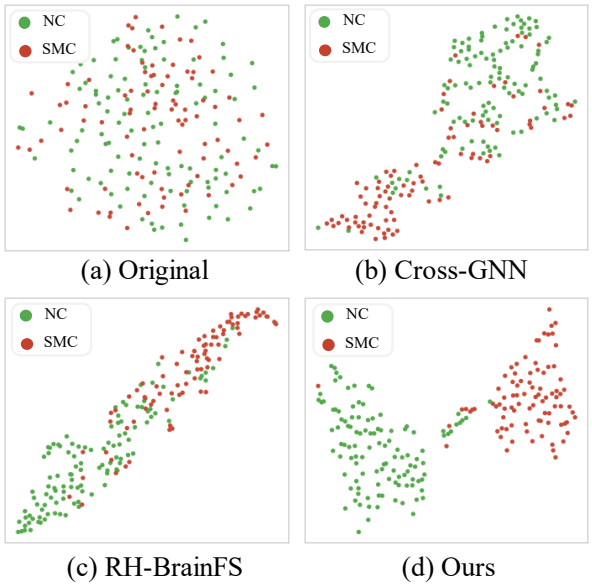


Fig. 5. T-SNE visualization for different methods on the NC vs. SMC task. Green circles represent the NC, while red circles represent SMC patients. The more closely the circles of the same color are clustered and the farther apart the circles of different colors are, the better the classification performance.

C. Decoupled Representation visualization

In this paper, we design a feature decoupling module to extract ineffective features, specific features and regional heterogeneous multi-modal features from FBN and SBN, respectively, and introduce the difference constraint to enhance the discrimination of these feature. To demonstrate the

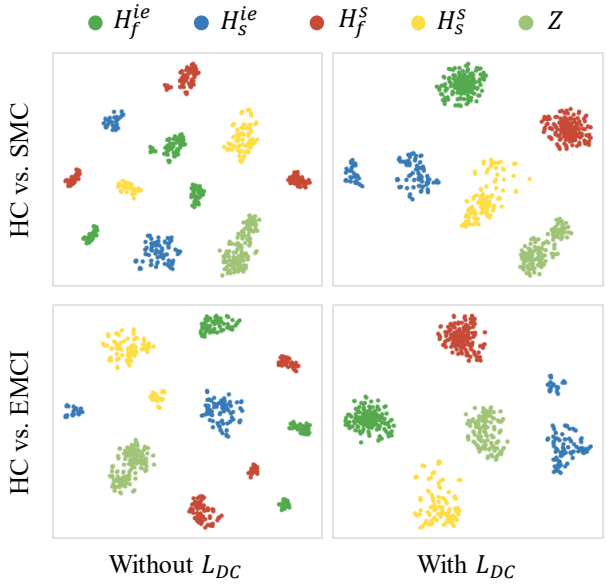


Fig. 6. Feature decoupling visualization. Circles of different colors represent different features. The closer the circles of the same color are clustered and the more dispersed the circles of different colors are, the better the decoupling performance.

effectiveness of decoupling, we illustrate the distribution of the aforementioned features before and after decoupling. As seen in Fig. 6, while FBN and SBN contain rich feature information, these features are entangled before decoupling, potentially interfering with the diagnosis of brain diseases. After applying feature decoupling, it not only aggregates features of the same kind but also extends the distance between different features. This indicates that the feature decoupling module can effectively reduce the interference of disease-unrelated information in modal-specific and multi-modal features, thus providing more valuable information for brain disease diagnosis.

D. Brain Network Visualization

To intuitively demonstrate the capability of different methods in capturing the interactions between FBN and SBN, we visualize the multi-modal brain networks constructed by the simple fusion method MKL, the interactive fusion method TSGR, and the proposed method in Fig. 7. To ensure consistency in numerical ranges, we normalize the brain networks. It can be seen from the figure that the brain networks constructed by MKL and TSGR exhibit an almost uniform distribution of strong and weak functional-structural couplings. This phenomenon suggests that both MKL and TSGR can

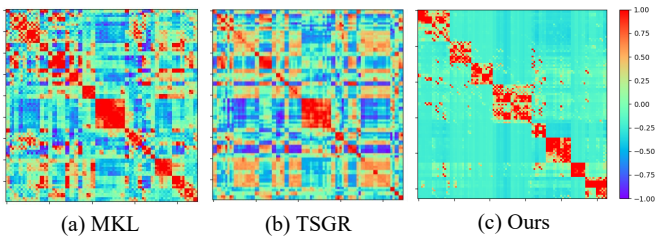
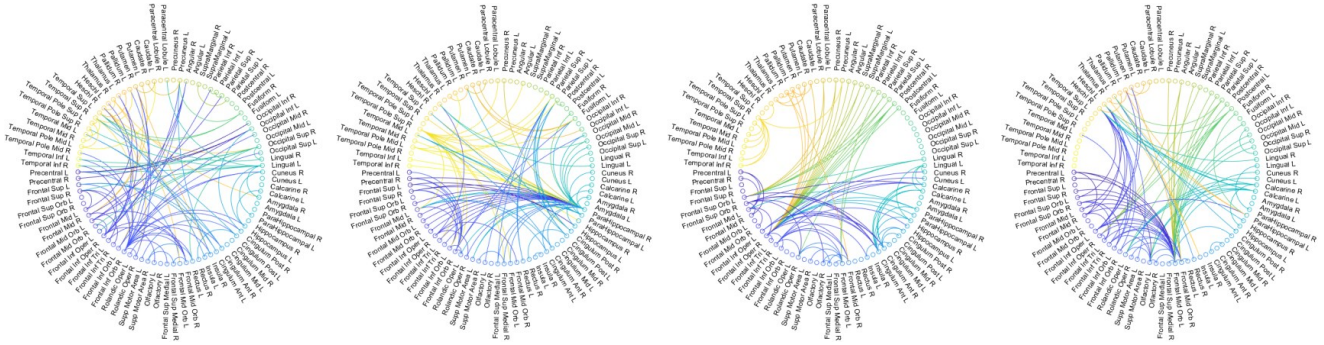


Fig. 7. Multi-modal brain networks constructed by different methods. Each element in the heatmap represents the connection strength between paired brain regions.



(a) NC vs. FLE (b) NC vs. TLE (c) NC vs. SMC (d) NC vs. EMCI

Fig. 8. The 100 most important Multi-modal brain connection on different classification tasks. Each node in the circular graph represents a brain region, and the lines between paired nodes indicate the presence of connections between these brain regions.

integrate multi-modal information to a certain extent, they treat the functional-structural coupling of all brain regions equally, making it difficult for them to effectively capture the intrinsic functional-structural interaction. Different from these methods, the proposed method can strengthen the interactions within subnetworks while weakening the interactions between subnetworks in the multi-modal brain network. This is because we simulate the hierarchical function-structural coupling between different brain regions through optimal transport, which comprehensively captures the regional heterogeneous interaction information in multi-modal brain networks.

E. Discriminative Brain Connections

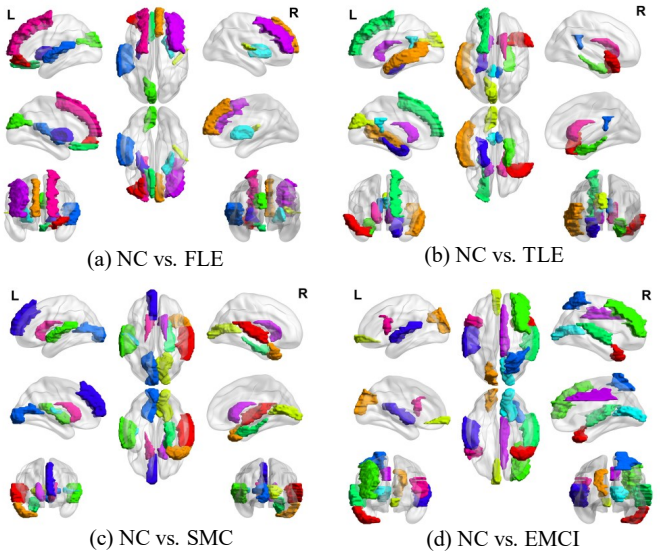
We further investigate the effectiveness of the proposed method in detecting biomarkers. For each diagnostic task, we use a t-test to identify the 100 most discriminative ($P < 0.05$) brain connections in multi-modal brain networks, as shown in Fig. 8. Specifically, for NC vs. FLE, discriminative brain connections are concentrated in the Precentral gyrus, Middle frontal gyrus, and Superior occipital gyrus. These regions are mainly responsible for motor control, cognitive regulation, and visual processing, which are closely related to FLE [48]. For NC vs. TLE, key brain connections focus on the Parahippocampal gyrus, Superior temporal gyrus, and Middle temporal gyrus, which are involved in memory and auditory processing [49]. TLE can cause memory deficits and auditory abnormalities, making these regions important biomarkers for distinguishing between HC and TLE patients. Additionally, we find some key connections outside the frontal and temporal lobes. This may be due to the abnormal electrical activity spreading through the brain network and affecting non-focal areas [50]. For the NC vs. SMC task, discriminative connections are concentrated in the Inferior frontal gyrus, Hippocampus, and Superior temporal gyrus. The identified brain connections correspond to regions associated with functional deficits in SMC patients, such as memory loss and blurred vision [51]. For NC vs. EMCI, the key brain connections are concentrated in the Superior frontal gyrus, Olfactory cortex, and Heschl gyrus, all of which have been proven to be related to cognitive impairment [52]. Therefore, the proposed method can effectively detect multi-modal biomarkers related to brain diseases.

F. Discriminative Brain Regions

To evaluate the proposed method's ability to identify important brain regions, we use t-test to detect the top 10 most significant ($P < 0.05$) brain regions for different diagnostic tasks. The experimental results are shown in Fig. 9. Specifically, for NC vs. FLE, the important brain regions include the Middle Frontal Gyrus, Frontal.Sup.L and Putamen. These brain regions play crucial roles in motor control and cognitive regulation and may serve as important markers for FLE [48]. For NC vs. TLE, key brain regions are concentrated in the Temporal.Pole.Sup.R, Temporal.Mid.L and Parahippocampal gyrus, which are typically involved in memory processing, language control and emotion perception, and are closely related to the occurrence of TLE [49]. For NC vs. SMC, the discriminative brain regions include the Superior Temporal Gyrus, Caudate Nucleus, and ParaHippocampal.R. Patients with SMC are usually accompanied by emotional anxiety and memory decline, which may be closely associated with abnormalities in the brain regions responsible for these functions [51]. For NC vs. EMCI, the important brain regions are concentrated in Temporal.Pole.Mid.R, Superior Temporal Gyrus and Frontal.Mid.R, which are responsible for semantic memory, functional execution and emotion regulation, respectively. Functional and structural abnormalities in these regions are key features in the development of EMCI [52]. Therefore, the proposed method can provide reasonable brain region markers for brain disease diagnosis.

VI. Conclusion

In this paper, we propose a regional heterogeneous multi-modal brain network analysis method based on optimal transport, and develop a feature decoupling module to eliminate disease-irrelevant multi-modal information. Specifically, we first divide the FBN and SBN into multiple subnetworks to obtain regional FBN and SBN. Then, we introduce penalty weights to reduce the communication cost between cross-modal brain regions within a subnetwork while increasing the cost between subnetworks to reflect the regional heterogeneity between FBN and SBN. Next, we apply optimal transport to simulate the transfer process of brain region hubness from FBN to SBN under regional heterogeneity constraints. This



(a) NC vs. FLE (b) NC vs. TLE

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

Fig. 9. Distribution of the 10 most important brain regions under different diagnostic tasks.

process not only fuses the features of multi-modal brain networks with regional heterogeneity, but also reveals the intrinsic functional-structural interactions of the brain. In addition, we design a feature decoupling module to suppress ineffective information and enhance the discrimination of specific and multi-modal features. Extensive experiments show that the proposed method can accurately diagnose temporal lobe epilepsy, frontal lobe epilepsy, significant memory concerns and early mild cognitive impairment, which proves the generalizability and reliability of the method. In future research, we will conduct validation on other types of brain disease datasets to further verify the robustness and broad applicability of the method.

REFERENCES

[1] Y. Zhang, P. Li, L. Cheng, M. Li, and H. Li, "Attention-based multiscale spatial-temporal convolutional network for motor imagery eeg decoding," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 2423–2434, 2024.

[2] M. A. Sayeed, S. P. Mohanty, E. Kougianos, and H. P. Zaveri, "eseiz: An edge-device for accurate seizure detection for smart healthcare," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 3, pp. 379–387, 2019.

[3] Q. Zhu, C. Zheng, Z. Zhang, W. Shao, and D. Zhang, "Dynamic confidence-aware multi-modal emotion recognition," *IEEE Transactions on Affective Computing*, vol. 15, no. 3, pp. 1358–1370, 2024.

[4] C. Zheng, W. Shao, D. Zhang, and Q. Zhu, "Prior-driven dynamic brain networks for multi-modal emotion recognition," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 389–398.

[5] A. Ashraf, Z. Qingjie, W. H. K. Bangyal, and M. Iqbal, "Analysis of brain imaging data for the detection of early age autism spectrum disorder using transfer learning approaches for internet of things," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 4478–4489, 2024.

[6] Q. Zhu, S. Li, X. Meng, Q. Xu, Z. Zhang, W. Shao, and D. Zhang, "Spatio-temporal graph hubness propagation model for dynamic brain network classification," *IEEE Transactions on Medical Imaging*, vol. 43, no. 6, pp. 2381–2394, 2024.

[7] D. Dai, J. Wang, J. Hua, and H. He, "Classification of adhd children through multimodal magnetic resonance imaging," *Frontiers in systems neuroscience*, vol. 6, p. 63, 2012.

[8] I. Sebenius, A. Campbell, S. E. Morgan, E. T. Bullmore, and P. Liò, "Multimodal graph coarsening for interpretable, mri-based brain graph neural network," in *International Workshop on Machine Learning for Signal Processing (MLSP)*, 2021, pp. 1–6.

[9] Q. Zhu, H. Wang, B. Xu, Z. Zhang, W. Shao, and D. Zhang, "Multimodal triplet attention network for brain disease diagnosis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3884–3894, 2022.

[10] Y. Yang, C. Ye, X. Guo, T. Wu, Y. Xiang, and T. Ma, "Mapping multimodal brain connectome for brain disorder diagnosis via cross-modal mutual learning," *IEEE Transactions on Medical Imaging*, vol. 43, no. 1, pp. 108–121, 2024.

[11] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J.-P. Thiran, R. Meuli, and P. Hagmann, "Predicting human resting-state functional connectivity from structural connectivity," *Proceedings of the National Academy of Sciences*, vol. 106, no. 6, pp. 2035–2040, 2009.

[12] K. Shen, B. Mišić, B. N. Cipollini, G. Bezgin, M. Buschkuehl, R. M. Hutchison, S. M. Jaeggi, E. Kross, S. J. Peltier, S. Everling *et al.*, "Stable long-range interhemispheric coordination is supported by direct anatomical projections," *Proceedings of the National Academy of Sciences*, vol. 112, no. 20, pp. 6473–6478, 2015.

[13] P. Fotiadis, L. Parkes, K. A. Davis, T. D. Satterthwaite, R. T. Shinohara, and D. S. Bassett, "Structure–function coupling in macroscale human brain networks," *Nature Reviews Neuroscience*, vol. 25, pp. 688–704, 2024.

[14] O. Sporns and R. F. Betzel, "Modular brain networks," *Annual review of psychology*, vol. 67, no. 1, pp. 613–640, 2016.

[15] Z.-Q. Liu, R. F. Betzel, and B. Misic, "Benchmarking functional connectivity by the structure and geometry of the human brain," *Network Neuroscience*, vol. 6, no. 4, pp. 937–949, 2022.

[16] J. Wu, Y. He, S. Liang, Z. Liu, J. Huang, W. Liu, J. Tao, L. Chen, C. C. Chan, and T. M. Lee, "Effects of computerized cognitive training on structure–function coupling and topology of multiple brain networks in people with mild cognitive impairment: a randomized controlled trial," *Alzheimer's Research & Therapy*, vol. 15, no. 1, p. 158, 2023.

[17] J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar *et al.*, "Functional network organization of the human brain," *Neuron*, vol. 72, no. 4, pp. 665–678, 2011.

[18] Y. Long, C. Chen, M. Deng, X. Huang, W. Tan, L. Zhang, Z. Fan, and Z. Liu, "Psychological resilience negatively correlates with resting-state brain network flexibility in young healthy adults: a dynamic functional magnetic resonance imaging study," *Annals of translational medicine*, vol. 7, no. 24, 2019.

[19] K. Ma, X. Wen, Q. Zhu, and D. Zhang, "Optimal transport based ordinal pattern tree kernel for brain disease diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 186–195.

[20] J. Dadashkarimi, A. Karbasi, and D. Scheinost, "Combining multiple atlases to estimate data-driven mappings between functional connectomes using optimal transport," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 386–395.

[21] B. Jiang, Y. Li, X. Wan, Y. Chen, Z. Tu, Y. Zhao, and J. Tang, "Mgdr: Multi-modal graph disentangled representation for brain disease prediction," in *Medical Image Computing and Computer Assisted Intervention*. Cham: Springer Nature Switzerland, 2024, pp. 302–312.

[22] H. Ye, Y. Zheng, Y. Li, K. Zhang, Y. Kong, and Y. Yuan, "Rh-brainfs: regional heterogeneous multimodal brain networks fusion strategy," *Advances in Neural Information Processing Systems*, vol. 36, pp. 59 286–59 303, 2024.

[23] R. Xu, Q. Zhu, S. Li, Z. Hou, W. Shao, and D. Zhang, "Mstgc: multi-channel spatio-temporal graph convolution network for multimodal brain networks fusion," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 2359–2369, 2023.

[24] E. Dhamala, D. S. Bassett, B. T. Yeo, and A. J. Holmes, "Functional brain networks are associated with both sex and gender in children," *Science Advances*, vol. 10, no. 28, p. eadn4202, 2024.

[25] M. Bianchini, M. Gori, and F. Scarselli, "Inside pagerank," *ACM Transactions on Internet Technology (TOIT)*, vol. 5, no. 1, pp. 92–128, 2005.

[26] W. Wang, L. Xiao, G. Qu, V. D. Calhoun, Y.-P. Wang, and X. Sun, "Multiview hyperedge-aware hypergraph embedding learning for multisite, multiatlas fmri based functional connectivity network analysis," *Medical Image Analysis*, vol. 94, p. 103144, 2024.

[27] J. Achterberg, D. Akarca, D. Strouse, J. Duncan, and D. E. Astle, "Spatially embedded recurrent neural networks reveal widespread links

between structural and functional neuroscience findings," *Nature Machine Intelligence*, vol. 5, no. 12, pp. 1369–1381, 2023.

[28] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," *Advances in neural information processing systems*, vol. 26, 2013.

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

[30] C. Tian, M. Zheng, W. Zuo, S. Zhang, Y. Zhang, and C.-W. Lin, "A cross transformer for image denoising," *Information Fusion*, vol. 102, p. 102043, 2024.

[31] C. Tian, Y. Yuan, S. Zhang, C.-W. Lin, W. Zuo, and D. Zhang, "Image super-resolution with an enhanced group convolutional neural network," *Neural Networks*, vol. 153, pp. 373–385, 2022.

[32] C. Tian, Y. Zhang, W. Zuo, C.-W. Lin, D. Zhang, and Y. Yuan, "A heterogeneous group cnn for image super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 5, pp. 6507–6519, 2024.

[33] E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng, and M. Joliot, "Automated anatomical labelling atlas 3," *Neuroimage*, vol. 206, p. 116189, 2020.

[34] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, "Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, 2017.

[35] T.-A. Song, S. R. Chowdhury, F. Yang, H. Jacobs, G. El Fakhri, Q. Li, K. Johnson, and J. Dutta, "Graph convolutional neural networks for alzheimer's disease classification," in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. IEEE, 2019, pp. 414–417.

[36] L. Huang, X. Ye, M. Yang, L. Pan, and S. hua Zheng, "Mnc-net: Multi-task graph structure learning based on node clustering for early parkinson's disease diagnosis," *Computers in Biology and Medicine*, vol. 152, p. 106308, 2023.

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

[38] X. Kan, W. Dai, H. Cui, Z. Zhang, Y. Guo, and C. Yang, "Brain network transformer," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25 586–25 599, 2022.

[39] X. Kan, H. Cui, J. Lukemire, Y. Guo, and C. Yang, "Fbnetgen: Task-aware gnn-based fmri analysis via functional brain network generation," in *International Conference on Medical Imaging with Deep Learning*. PMLR, 2022, pp. 618–637.

[40] S. Dahan, L. Z. Williams, D. Rueckert, and E. C. Robinson, "Improving phenotype prediction using long-range spatio-temporal dynamics of functional connectivity," in *Machine Learning in Clinical Neuroimaging: 4th International Workshop, MLCN 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 4*. Springer, 2021, pp. 145–154.

[41] D. Chen, M. Liu, Z. Shen, X. Zhao, Q. Wang, and L. Zhang, "Learnable subdivision graph neural network for functional brain network analysis and interpretable cognitive disorder diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 56–66.

[42] J. R. San Cristóbal Mateo and J. R. S. C. Mateo, "Weighted sum method and weighted product method," *Multi Criteria Analysis in the Renewable Energy Industry*, pp. 19–22, 2012.

[43] K.-H. Shih, C.-T. Chiu, J.-A. Lin, and Y.-Y. Bu, "Real-time object detection with reduced region proposal network via multi-feature concatenation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 6, pp. 2164–2173, 2019.

[44] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, A. D. N. Initiative *et al.*, "Multimodal classification of alzheimer's disease and mild cognitive impairment," *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.

[45] X. Wang, M. Zhu, D. Bo, P. Cui, C. Shi, and J. Pei, "Am-gcn: Adaptive multi-channel graph convolutional networks," in *Proceedings of the 26th ACM SIGKDD International conference on knowledge discovery & data mining*, 2020, pp. 1243–1253.

[46] Y. Kang, R. Wang, E. Shi, J. Wu, S. Yu, and S. Zhang, "Exploring brain function-structure connectome skeleton via self-supervised graph-transformer approach," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 308–317.

[47] C. Tian, M. Zheng, B. Li, Y. Zhang, S. Zhang, and D. Zhang, "Perceptive self-supervised learning network for noisy image watermark removal," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 8, pp. 7069–7079, 2024.

[48] W. Liu, Q. Yue, Q. Gong, D. Zhou, and X. Wu, "Regional and remote connectivity patterns in focal extratemporal lobe epilepsy," *Annals of Translational Medicine*, vol. 9, no. 14, 2021.

[49] M. Petrides, "On the evolution of polysensory superior temporal sulcus and middle temporal gyrus: a key component of the semantic system in the human brain," *Journal of Comparative Neurology*, vol. 531, no. 18, pp. 1987–1995, 2023.

[50] C. Yang, Z. Liu, Q. Wang, Q. Wang, Z. Liu, and G. Luan, "Epilepsy as a dynamical disorder orchestrated by epileptogenic zone: a review," *Nonlinear Dynamics*, vol. 104, pp. 1901–1916, 2021.

[51] S. H. Kang, Y. H. Park, J. P. Kim, J.-S. Kim, C. H. Kim, H. Jang, H. J. Kim, S.-B. Koh, D. L. Na, J. Chin *et al.*, "Cortical neuroanatomical changes related to specific neuropsychological deficits in subcortical vascular cognitive impairment," *NeuroImage: Clinical*, vol. 30, p. 102685, 2021.

[52] L. Du, F. Liu, K. Liu, X. Yao, S. L. Risacher, J. Han, A. J. Saykin, and L. Shen, "Associating multi-modal brain imaging phenotypes and genetic risk factors via a dirty multi-task learning method," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3416–3428, 2020.