

Lateralization-aware multi-view high-order graph learning for brain disorder classification

Jiazhen Ye ^a, Manman Yuan ^{a,*}, Yan Zhao ^b, Can Yin ^a, Mengyi Shao ^a,
Jürgen Kurths ^{c,d}

^a School of Computer Science, Inner Mongolia University, Hohhot, 010021, China

^b Department of Neurology, Inner Mongolia People's Hospital, Hohhot, 010010, China

^c Institute of Physics, Humboldt University of Berlin, Berlin, 12489, Germany

^d Potsdam Institute for Climate Impact Research, Potsdam, 14473, Germany

ARTICLE INFO

Keywords:

Brain networks learning
Graph neural networks
Multi-view learning
Lateralization modeling
High-order interactions
Brain disorder classification

ABSTRACT

Brain networks have emerged as a crucial tool for exploring brain organization and aiding in brain disorder classification. Recent graph neural network (GNN)-based methods have demonstrated strong potential for brain network analysis. However, most of these methods insufficiently model lateralized heterogeneity induced by hemispheric asymmetry. Additionally, traditional message-passing schemes mainly rely on direct neighbors, which restricts their ability to capture high-order synergistic interactions among distant brain regions. To overcome these challenges, this paper proposes a Lateralization-Aware Multi-View High-Order Graph Learning (LMHGL) framework for brain disorder classification, which provides a unified scheme to model hemispheric asymmetry, supporting high-order intra-hemispheric interactions and multi-view collaboration at the decision level. Specifically, LMHGL introduces an anatomically constrained heat diffusion mechanism to model high-order intra-hemispheric interactions. By propagating information along structural connectivity, it captures biologically plausible multi-hop dependencies and enhances the representation of lateralized brain organization. Furthermore, a multi-view collaborative decision mechanism integrates global and hemispheric-specific features, enabling the joint characterization of holistic and lateralized pathological patterns. Extensive experiments on three real-world brain disorder datasets demonstrate that LMHGL consistently outperforms state-of-the-art methods, highlighting the effectiveness of jointly modeling hemispheric asymmetry and anatomically constrained high-order interactions for learning discriminative brain network representations.

1. Introduction

Recent developments in neuroimaging technologies have greatly enhanced our understanding of the brain's functional and structural organization. For example, functional magnetic resonance imaging (fMRI) reveals functional activity within specific anatomical brain regions of interest (ROIs) by measuring blood oxygenation level-dependent (BOLD) signals, while diffusion tensor imaging (DTI) illustrates structural connectivity via white-matter fiber tracts between ROIs. A key instrument for integrating these neuroimaging observations is the brain network, which represents macroscopic brain organization by mapping ROIs as nodes and their connections as edges. Then, graph-theoretic tools can be applied to these brain networks, allowing for the quantification and exploration of the organization and interactions among brain regions (Luo et al., 2024). Consequently, brain networks have been extensively used to investigate brain organization and

characterize neurological and psychiatric disorders (Fang et al., 2025; Li et al., 2021; Liu et al., 2024; Wen et al., 2022; Yuan et al., 2024; Zhu et al., 2024).

In recent years, graph neural networks (GNNs) (Gilmer et al., 2017; Kipf & Welling, 2017; Velickovic et al., 2018) have emerged as a powerful tool for brain network analysis, attributed to their ability to model complex relational structures in graph-structured data. Through the message-passing mechanism (Gilmer et al., 2017), GNNs iteratively aggregate and transform information from neighboring nodes, enabling each ROI to integrate contextual information from its connected regions. This has led to successful application in various brain network analysis scenarios (Luo et al., 2024), such as static, dynamic, or multi-modal settings, offering comprehensive and diverse approaches to understand brain networks. Despite these advances, there are still several fundamental challenges in effectively modeling the complex organization of brain networks.

* Corresponding author at: No. 235 West College Road, Saihan District, Hohhot, Inner Mongolia, PR China.

E-mail address: yuanman@imu.edu.cn (M. Yuan).

<https://doi.org/10.1016/j.eswa.2026.132999>

Received 10 March 2026; Received in revised form 30 April 2026; Accepted 22 May 2026

Available online 27 May 2026

0957-4174/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

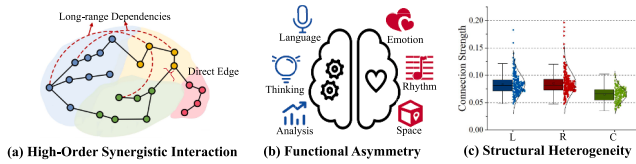


Fig. 1. (a) High-order synergistic interactions among distant brain regions beyond direct connections. (b) Functional hemispheric asymmetry in the human brain. (c) Differences in connection strengths for left-, right-, and cross-hemispheric (L, R, and C) edges, estimated from DTI-derived fractional anisotropy (FA) and observed in our experimental datasets.

A fundamental limitation of existing GNN-based methods lies in their reliance on localized message passing, which primarily aggregates information from direct neighbors, thereby restricting their ability to capture high-order synergistic interactions among distant brain regions. Such interactions reflect long-range dependencies among ROIs that are not directly connected, but exhibit coordinated functional behavior (Yu et al., 2024), as illustrated in Fig. 1(a). To address this issue, numerous studies have investigated hypergraph-based formulations (Han et al., 2026; Hao et al., 2024; Yuan et al., 2024), in which groups of ROIs are organized into hyperedges to approximate high-order interactions. However, these hyperedges are typically derived from pairwise correlations, causing the resulting representations to remain largely dependent on low-order statistics. Alternatively, attention-based models (Kan et al., 2022; Pei et al., 2025; Xue et al., 2025b; Yu et al., 2024) attempt to capture long-range dependencies by dynamically weighting distant nodes, yet such data-driven mechanisms often lack anatomical constraints and may lead to biologically implausible interactions. Therefore, developing a principled approach for modeling high-order interactions is still an open problem in brain network analysis.

In addition, another important challenge stems from the insufficient and unstructured modeling of lateralized heterogeneity in brain networks. In neuroscience, hemispheric asymmetry indicates systematic functional and structural differences between the two hemispheres (Toga & Thompson, 2003; Wu et al., 2022). Functionally, language and logical reasoning are mainly left-lateralized (Thiebaut de Schotten et al., 2011), while spatial and affective processing are usually right-lateralized (Fernández-Carriba et al., 2002), as depicted in Fig. 1(b). Structurally, hemispheric asymmetry is reflected in differences in gray matter volume and white matter connectivity. Our quantitative analysis based on DTI in Fig. 1(c) further demonstrates that intra-hemispheric connections exhibit relatively consistent strengths, while inter-hemispheric connections display greater variability across datasets, indicating intrinsic topological heterogeneity induced by hemispheric asymmetry. Despite these observations, most existing GNN-based methods propagate information uniformly across the entire graph, overlooking such biologically meaningful asymmetries. Including static (Han et al., 2026; Wen et al., 2022; Yuan et al., 2024), dynamic (Fang et al., 2025; Liu et al., 2024), and multi-modal frameworks (Hemker et al., 2024; Jiang et al., 2025; Lee et al., 2024), which typically operate on whole-brain representations without incorporating lateralized priors. More recently, hemispheric-aware modeling strategies (Gu et al., 2024; Shi et al., 2025; Ye et al., 2025) have been developed to model lateralization by separate hemispheric components. However, they remain limited in two aspects: First, the key intra-hemispheric modeling still relies on standard message passing, overlooking the high-order interactions among brain regions; Second, hemispheric representations are typically collapsed into a unified global embedding for downstream tasks, thereby weakening their independence and limiting the exploitation of hemispheric-specific inductive biases.

To address the above challenges, we propose LMHGL, a Lateralization-Aware Multi-view High-Order Graph Learning framework for brain disorder classification. The key idea is to model hemispheric asymmetry within a unified framework that captures

high-order intra-hemispheric interactions while enabling decision-level multi-view collaboration by leveraging complementary information and lateralized pathological patterns. To this end, a Tripartite Network Deconstruction Strategy (TNDS) is first introduced, based on hemispheric asymmetry, to partition each brain network into left-, right-, and cross-hemispheric subnetworks as separate views, thereby providing a structural basis for lateralization-aware modeling. These subnetworks are then processed by a Multi-View Heat-Assisted Graph Convolutional Network (MHGCN) to learn subnetwork-specific embeddings and a global brain representation by adaptively multi-view fusion. Within this module, an anatomically constrained heat diffusion is selectively applied to left/right-hemispheric subnetworks to model high-order interactions via multi-hop information propagation along structural connectivity, while the cross-hemispheric subnetwork preserves direct inter-hemispheric coupling patterns. Furthermore, a Multi-View Decision Mechanism (MVDM) is designed to jointly incorporate global and hemispheric-specific representations at the decision level, enabling the model to leverage complementary multi-view information and lateralized pathological patterns for refined prediction. Extensive experiments on three public real-world brain disorder datasets demonstrate the effectiveness of the proposed LMHGL framework, highlighting its superiority in modeling hemispheric asymmetry and anatomically constrained high-order interactions for learning discriminative brain network representations.

The main contributions of this work are threefold:

1. We propose LMHGL, a unified framework to model hemispheric asymmetry. It captures high-order intra-hemispheric interactions and enables decision-level multi-view collaboration to leverage complementary and lateralized pathological patterns.
2. We introduce an anatomically constrained heat diffusion to model high-order interactions. It propagates information along structural connectivity, enabling the modeling of biologically plausible multi-hop dependencies.
3. Extensive experiments on three real-world datasets demonstrate the effectiveness of LMHGL. The results further show its superiority in modeling hemispheric asymmetry and high-order interactions.

The rest of this paper is organized as follows. Section 2 reviews related work, Section 3 outlines the proposed methodology, Section 4 details the experimental setup, Section 5 reports and discusses the results on the real-world data, Section 6 discusses the limitations and future research directions, and Section 7 concludes the paper.

2. Related work

2.1. GNN-based brain network

In recent years, GNNs have emerged as a promising tool for brain network analysis, attributed to their ability to capture complex relational patterns in graph-structured data. Current methods can be broadly categorized into three types: static, dynamic, and multi-modal. Static models usually operate on whole-brain networks in specific states and learn subject-level features using graph convolutional encoders. Representative studies such as Li et al. (2021), Qiu et al. (2024), Thapaliya et al. (2025), Yuan et al. (2024) adhere to this paradigm by constructing a single connectivity graph per subject and conducting graph-level inference for disease classification. Dynamic models focus on capturing temporal variations in connectivity by constructing time-resolved graphs over sliding windows. For instance, Fang et al. (2025) presented an interpretable spatio-temporal GNN that exploits information-bottleneck constraints to extract task-relevant dynamics, while Liu et al. (2024) developed a spatio-temporal multi-hypergraph framework to model transient connectivity states. Multi-modal models aim to integrate complementary information from multiple imaging modalities by learning modality-specific representations. For example, HEALNet (Hemker

et al., 2024) extracts heterogeneous features from each modality using dedicated encoders, while several dual-branch architectures (Jiang et al., 2025; Lee et al., 2024) jointly model fMRI-based functional graphs and DTI-derived structural graphs, allowing the network to exploit cross-modality complementarities. Nevertheless, these methods often treat the brain as a homogeneous graph and overlook the intrinsic lateralized heterogeneity induced by hemispheric asymmetry, which may constrain their ability to learn disorder-related pathological patterns.

2.2. Hemispheric asymmetry in brain networks

Hemispheric asymmetry refers to the structural and functional differences influenced by evolutionary, experiential, and pathological factors (Toga & Thompson, 2003). These asymmetry patterns are dynamic and exhibit characteristic disruptions in neuropsychiatric disorders. For instance, Alzheimer's disease typically presents a left-dominant vulnerability, with pathological progression predominantly affecting the left hemisphere (Ye et al., 2025). Schizophrenia involves right-lateralized disruptions in network efficiency, while autism spectrum disorder exhibits atypical lateralization, such as right-shifted or attenuated language asymmetry (Sun et al., 2017). Such deviations provide critical insights into their pathophysiology. Drawing on this, recent methods (Gu et al., 2024; Shi et al., 2025; Ye et al., 2025) have begun to develop hemispheric-aware modeling strategies. For example, LG-DBGL learns separate representations for hemispheric subnetworks and dynamically fuses them into a global representation for Alzheimer's disease identification. HeBrainGNN (Shi et al., 2025) aligns inter- and intra-hemispheric representations to enhance structural encoding. FC-HGNN (Gu et al., 2024) first encodes intra- and inter-hemispheric connections to learn individual-level embedding, and then constructs a heterogeneous population graph to improve mental disorder identification. Nevertheless, these methods remain limited in fully exploiting lateralized information. First, intra-hemispheric modeling is limited by standard message-passing, overlooking the high-order interactions among distributed brain regions. Second, hemispheric representations are typically collapsed into a unified global embedding for the downstream task, limiting the effective utilization of hemispheric-specific pathological patterns. In contrast, our approach addresses these limitations within a unified framework by modeling intra-hemispheric high-order synergistic interactions and preserving hemispheric-specific representations for decision-level collaboration.

2.3. High-order modeling in brain networks

High-order interactions in brain networks reflect complex, indirect dependencies among brain regions beyond direct pairwise connections. Modeling such interactions is crucial for capturing long-range, coordinated neural activity and understanding distributed brain functions. Existing approaches for high-order modeling can be broadly categorized into two paradigms: hypergraph-based and attention-based methods. Hypergraph-based methods model high-order interactions by constructing hyperedges that connect multiple nodes. For example, MHSA (Yuan et al., 2024) builds hyperedges based on top-k strongest connections derived from spatio-temporal fused fMRI features. WHGCN (Hao et al., 2024) constructs K-nearest neighbor hypergraphs from sMRI data to capture inter-subject relationships over time. HGST (Han et al., 2026) leverages sparse representation to identify relevant regions and employs self-supervised learning to extract high-order features. However, these methods typically construct hyperedges based on pairwise correlations or heuristic strategies, resulting in representations that are still fundamentally derived from low-order statistics and lack a principled mechanism to capture genuine multi-hop dependencies. Furthermore, attention-based methods offer an alternative by modeling long-range dependencies through learnable attention weights. For instance, BrainNetTF (Kan et al., 2022) applies multi-head self-attention over ROI connectivity features, while subsequent works such as ALTER (Yu et al.,

2024), LHDFormer (Xue et al., 2025b), and CAGT (Pei et al., 2025) further incorporate random walk sampling, temporal modeling, or multi-scale feature fusion to enhance long-range interaction modeling. Nevertheless, these approaches are largely data-driven and lack explicit anatomical constraints, which may lead to spurious or biologically implausible interactions. In contrast, high-order interactions in our methods are modeled by heat diffusion over the structural connectome, enabling principled multi-hop information propagation while preserving biological plausibility.

3. Methodology

3.1. Overview

In this section, we present the proposed LMHGL framework for brain disorder classification. An overview of the framework is illustrated in Fig. 2, which consists of three components. First, a Tripartite Network Deconstruction Strategy (TNDS) decomposes the brain network into hemispheric subnetworks, which establishes a structural basis for lateralization-aware modeling. Second, a Multi-View Heat-Assisted Graph Convolutional Network (MHGCN) learns subnetwork-specific representations and integrates them into a global brain representation, where anatomically constrained heat diffusion is used to model high-order intra-hemispheric interactions. Third, a Multi-View Decision Mechanism (MVD) jointly leverages subnetwork-specific and global representations at the decision level to enable collaborative prediction.

Notably, LMHGL differs from existing methods in both multi-view and diffusion-based modeling. Firstly, most multi-view methods construct views based on multi-modal (Hemker et al., 2024; Jiang et al., 2025; Lee et al., 2024) or temporal dynamic information (Fang et al., 2025; Liu et al., 2024), and perform fusion at the feature level. Our multi-view setting originates from the subnetwork views based on the hemispheric anatomical structure constructed by TNDS. These subnetworks describe potential lateralized pathological patterns and are directly involved in decision-making through MVD, enabling the effective utilization of hemisphere-specific abnormalities. Secondly, existing methods (Sim et al., 2024; Yu et al., 2024) that use heat diffusion or related mechanisms often perform data-driven propagation over entire brain network. Such designs lack anatomical constraints and may lead to excessive cross-hemispheric smoothing, thereby weakening inherent lateralized characteristics. In contrast, the heat diffusion in MHGCN is anatomically guided, modeling multi-hop information propagation along structural connectivity. By treating each hemisphere as a functionally coherent unit, the diffusion process is primarily confined within hemispheres, effectively preserving lateralized patterns.

In summary, LMHGL constitutes a unified modeling framework tailored to lateralized brain networks, based on anatomically grounded subnetwork construction, which supports constrained high-order interaction modeling and decision-level exploitation of hemispheric-specific patterns.

3.2. Preliminary

Subject-specific brain networks are constructed by integrating resting-state fMRI and DTI data under a standardized AAL atlas (Tzourio-Mazoyer et al., 2002). Each brain network is modeled as a weighted graph $G = (V, E)$, where $V = \{v_i\}_{i=1}^N$ denotes the set of N nodes corresponding to anatomical brain ROIs, and $E = \{e_{i,j} \mid i, j \in V\}$ denotes the set of edges capturing the structural connections derived by DTI. Each node $v_i \in V$ is linked to a feature vector $X_i \in \mathbb{R}^T$ derived from the BOLD time series of the corresponding ROI, forming the node feature matrix $X \in \mathbb{R}^{N \times T}$ with a length of T time points. Additionally, structural connectivity is represented by the adjacency matrix $A^s \in \mathbb{R}^{N \times N}$, where $A^s_{i,j} \in [0, 1]$ indicates the normalized Fractional Anisotropy (FA) value of the white-matter tract connecting ROIs v_i and v_j . A greater FA

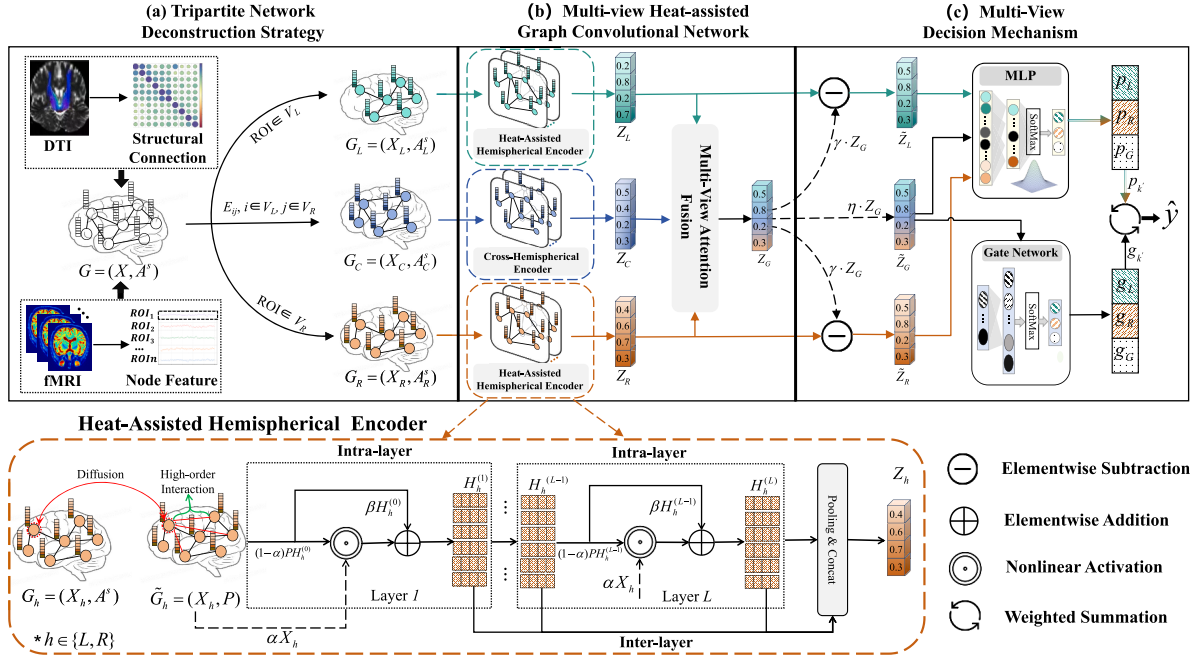


Fig. 2. Overview of the proposed LMHGL framework. (a) Tripartite network decomposition strategy (TNDS) partitions each subject's brain network G into left-hemispheric (G_L), right-hemispheric (G_R), and cross-hemispheric (G_C) subnetworks. (b) Multi-view heat-assisted graph convolutional network (MHGCN) encodes the G_L and G_R via a heat-assisted hemispherical encoder and the G_C via a cross-hemispherical encoder to generate view-specific embeddings Z_L , Z_R and Z_C , respectively, followed by multi-view attention fusion to produce a global brain representation Z_G . (c) Multi-view decision mechanism (MVDMM) decorrelates Z_L and Z_R from Z_G to obtain $\tilde{Z}_L$, $\tilde{Z}_R$, and $\tilde{Z}_G$, and then adaptively weights their respective predictions p_L , p_R , and p_G using gating weights g_L , g_R , and g_G to obtain the final class probability $\hat{y}$.

value indicates stronger directional diffusion along fiber tracts, indicating more reliable structural connectivity. Thus, each brain network can be compactly denoted as $G = (X, A^s)$, offering a unified representation of both functional activity and anatomical connectivity for subsequent graph-based modeling.

3.3. Tripartite network deconstruction strategy

The brain displays pronounced hemispheric asymmetry, characterized by lateralized functional specialization and distinct intra- and inter-hemispheric connectivity patterns, as shown in Fig. 1(b)-(c). To disentangle these heterogeneous structures, we build upon the strategies described in Gu et al. (2024), Ye et al. (2025) and formulate this idea as TNDS, which partitions each brain network G into three anatomically grounded and complementary subnetwork views: left-hemispheric G_L , right-hemispheric G_R , and cross-hemispheric G_C .

Specifically, in line with the AAL atlas convention, ROIs are indexed such that even-numbered nodes correspond to the left hemisphere while odd-numbered nodes correspond to the right. Thus, the node set $V = \{v_i\}_{i=1}^N$ can be split into two disjoint hemispheric subsets:

$$\mathcal{V}_h = \begin{cases} \{v_{2k} \mid k = 1, \dots, N/2\}, & \text{if } h = L, \\ \{v_{2k-1} \mid k = 1, \dots, N/2\}, & \text{if } h = R, \end{cases} \quad (1)$$

with $V_L \cap V_R = \emptyset$ and $V_L \cup V_R = V$. For each hemisphere $h \in \{L, R\}$, an intra-hemispheric subnetwork view $G_h = (X_h, A_h^s)$ is constructed by extracting the corresponding node features and structural connections. An index mapping $\phi_h(k) = 2k - \delta_h$ is used with $\delta_L = 0$ and $\delta_R = 1$, such that:

$$A_h^s[k, l] = A^s[\phi_h(k), \phi_h(l)], \quad X_h[k, :] = X[\phi_h(k), :]. \quad (2)$$

To model inter-hemispheric coupling, a cross-hemispheric subnetwork view $G_C = (X_C, A_C^s)$ is further constructed over the full node set, where the adjacency matrix A_C^s includes only cross-hemispheric struc-

tural connections:

$$A_C^s[i, j] = \begin{cases} A^s[i, j], & \text{if } i \in V_L, j \in V_R \text{ or } i \in V_R, j \in V_L, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

and node feature matrix $X_C \in \mathbb{R}^{N \times T}$ is defined as:

$$X_C[i, :] = \begin{cases} X[i, :], & \text{if } \sum_j A_C^s[i, j] > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where zero padding is used to maintain consistent node alignment across views.

Through this decomposition, TNDS provides a structured basis for subsequent multi-view modeling of lateralized brain networks. Notably, this tripartite decomposition incurs negligible computational overhead, as it only involves simple slicing operations, and each resulting subnetwork is defined on a smaller or sparser connectivity structure.

3.4. Multi-view heat-assisted graph convolutional network

This subsection focuses on learning embedding from the subnetwork views G_L , G_R , and G_C , each capturing distinct functional roles in the brain. Specifically, the heat-assisted hemispherical encoder is applied to G_L and G_R to model intra-hemispheric high-order interactions, while a cross-hemispheric encoder captures inter-hemispheric coupling patterns from G_C . The resulting view-specific embeddings are further integrated into a global brain representation via a multi-view attention fusion module.

Heat-Assisted Hemispherical Encoder. Modeling high-order interactions in brain networks is essential yet challenging, as conventional GNNs are inherently limited to localized message passing along direct edges. Existing high-order approaches, such as hypergraph-based and attention-based methods, either rely on constructions derived from low-order statistics or introduce unconstrained long-range interactions that may lack biological plausibility (see Section 2.3). To address these limitations, we model high-order interactions via anatomically constrained

heat diffusion over the structural connectome, thereby capturing continuous multi-hop dependencies consistent with potential anatomical structures. Moreover, prior neuroimaging studies (Abdelnour et al., 2014) have shown that diffusion processes over structural connectivity can approximate functional coupling patterns, offering biological support for this design. Consistent with hemispheric specialization, the diffusion process is applied separately within each hemisphere.

Formally, for each hemisphere $h \in \{L, R\}$, let $G_h = (X_h, A_h^s)$ be the corresponding subnetwork. The diffusion is defined on the normalized graph Laplacian $\hat{L}_h = I_n - D_h^{-1/2} A_h^s D_h^{-1/2}$, where $(D_h)_{i,i} = \sum_j (A_h)_{i,j}$ is the degree matrix. Mathematically, the heat kernel is governed by:

$$P_h = e^{-t\hat{L}_h} = \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} \hat{L}_h^k, \quad (5)$$

where $t \in \mathbb{R}^+$ controls the diffusion scale. Each term $\hat{L}_h^k$ aggregates information from k -hop neighbors, thus encoding high-order dependencies, while the exponential weighting $\frac{(-t)^k}{k!}$ diminishes the influence of distant nodes to preserve that local geometric structure (Kondor, 2002). Compared with hypergraph and attention-based methods, the proposed formulation defines high-order interactions directly through the graph Laplacian and constrains information propagation along structural connectivity (as encoded in A_h^s), resulting in a principled and anatomically grounded representation.

Then, the diffused graph $\tilde{G}_h = (X_h, P_h)$ is utilized to guide message propagation in an intra-inter GCN architecture inspired by Chen et al. (2020), Xu et al. (2018), which aims to reduce over-smoothing and stabilize deep graph representations. The intra-layer propagation at the l th layer is defined as:

$$H_h^{(l)} = \text{ReLU}((1 - \alpha)P_h H_h^{(l-1)} + \alpha X_h) + \beta H_h^{(l-1)}, \quad (6)$$

where $H_h^{(l)}$ refers to the node embedding matrix, and $\alpha, \beta \in [0, 1]$ control the retention of initial node features and the strength of residual connections, respectively. To capture hierarchical representations across varying propagation depths, the outputs of all layers are aggregated through inter-layer fusion:

$$Z_h = \text{Mean}(\text{Linear}([H_h^{(1)} | \dots | H_h^{(L)}])), \quad (7)$$

where $[\cdot | \cdot]$ indicates feature concatenation, $\text{Linear}(\cdot)$ denotes a learnable linear projection, and $\text{Mean}(\cdot)$ signifies average pooling across nodes.

Cross-Hemispheric Encoder. The cross-hemispheric subnetwork $G_C = (X_C, A_C^s)$ captures only inter-hemispheric structural connections and thus primarily captures direct coupling patterns between the two hemispheres. Since such interactions are direct connections and do not necessitate further high-order modeling, we adopt the Graph Isomorphism Network (GIN) (Xu et al., 2019) as the backbone encoder, leveraging its strong expressive capability for characterizing local structures. The update rule at the l th layer is expressed as:

$$H_C^{(l)} = \sigma((\epsilon^{(l)} \cdot I_n + A_C^s) H_C^{(l-1)} W_C^{(l)}), \quad (8)$$

where $H_C^{(0)} = X_C$, $\epsilon^{(l)} \in \mathbb{R}^+$ is a learnable scalar for balancing self-information and neighborhood aggregation, $W_C^{(l)} \in \mathbb{R}^{d_{l-1} \times d_l}$ is a trainable weight matrix, and $\hat{A}_C^s = I_n - D_C^{-1/2} A_C^s D_C^{-1/2}$ is the normalized adjacency matrix. After L layers of propagation, the final cross-hemispheric representation is expressed through mean pooling:

$$Z_C = \text{Mean}(H_C^{(L)}). \quad (9)$$

Multi-View Attention Fusion. Given the hemispheric-specific embeddings $\{Z_k\}_{k \in \mathcal{K}}$ with $\mathcal{K} = \{L, R, C\}$, we construct a global brain representation by adaptively aggregating embeddings across subnetworks. Specifically, we design a lightweight attention-based fusion strategy that assigns importance scores based on the overall activation strength of each norm-embedding. For each $k \in \mathcal{K}$, the attention score $a_k \in \mathbb{R}^+$ is computed as:

$$a_k = \frac{1}{\sqrt{d}} \text{MLP}(\|Z_k\|_2), \quad (10)$$

where $\|Z_k\|_2$ denotes the L_2 norm of the embedding, the multi-layer perceptron $\text{MLP}(\cdot)$ maps scalar norm to a learnable attention score, and the factor $1/\sqrt{d}$ stabilizes gradient scaling. This design is motivated by two considerations. First, the L_2 norm of the embedding provides a compact proxy for its overall activation strength, which correlates with its information content and discriminative capacity. Second, the lightweight MLP module avoids introducing excessive parameters, thereby reducing the risk of overfitting in limited-sample clinical settings.

Subsequently, these attention coefficients are normalized to make their sum equal one, allowing for a probabilistic interpretation of each view's contribution:

$$\beta_k = \frac{\exp(a_k)}{\sum_{k \in \mathcal{K}} \exp(a_k)}. \quad (11)$$

Finally, the global brain representation Z_G is derived as a weighted summation of each hemispheric-specific embedding:

$$Z_G = \sum_{k \in \mathcal{K}} \beta_k Z_k. \quad (12)$$

3.5. Multi-view decision mechanism

After the encoding process outlined previously, subnetwork embeddings Z_L , Z_R , and Z_C , as well as the global brain network representation Z_G are obtained. It is noteworthy that Z_C is not used as an independent decision input because its information has already been explicitly incorporated into Z_G via the multi-view attention fusion. Conversely, Z_L and Z_R are preserved, as they may retain hemispheric-specific and lateralization-related disease patterns that could be diminished in the global representation. In this stage, collaborative decision-making is performed by jointly exploiting Z_L , Z_R , and Z_G . Nevertheless, since these representations share overlapping information, directly combining their predictions could lead to redundancy and obscure respective contributions. Consequently, this paper first applies a decorrelation preprocessing step filtering that suppresses global information in Z_L and Z_R shared with Z_G , to promote respective representational independence:

$$\tilde{Z}_L = Z_L - \gamma \cdot Z_G, \quad (13)$$

$$\tilde{Z}_R = Z_R - \gamma \cdot Z_G, \quad (14)$$

$$\tilde{Z}_G = \eta \cdot Z_G, \quad (15)$$

where $\gamma, \eta \in (0, 1)$ are scaling factors that determine the degree of global information suppression and the emphasis on global features, respectively. This operation enables $\tilde{Z}_L$ and $\tilde{Z}_R$ to focus on lateralized patterns that are not fully captured by the global representation while maintaining a scaled global embedding for holistic decision-making. Then, each decorrelated embedding is fed into a shared MLP followed by a Softmax layer to generate the corresponding class probability distribution:

$$p_{k'} = \text{softmax}(\text{MLP}(\tilde{Z}_{k'})), k' \in \{L, R, G\}. \quad (16)$$

These prediction results provide diagnostic insights from different perspectives. To facilitate collaborative decision-making, they are fed into a gating network to learn the relative contributions of each prediction. Specifically, the gating vector $g \in \mathbb{R}^3$ is derived from the global representation $\tilde{Z}_G$:

$$g = \text{softmax}(\text{Tanh}(W_2 \cdot \text{ReLU}(W_1 \cdot \tilde{Z}_G))/\tau), \quad (17)$$

where W_1 and W_2 are learnable projection matrices, and τ is a temperature parameter that controls the sharpness of the gating distribution. The final prediction $\hat{y}$ is obtained by applying the gated vector weighted class probability:

$$\hat{y} = \sum_{k' \in \{L, R, G\}} g_{k'} \cdot p_{k'}. \quad (18)$$

Finally, $\hat{y}$ is assessed against the ground-truth label y using cross-entropy loss for end-to-end training.

Table 1
Demographic and clinical characteristics of subjects.

Dataset	Group	Number	Gender		Age (Mean/Std)
			Male	Female	
ADNI	Normal Control	115	85	30	88/5.1
	Mild Cognitive Impairment	51	33	18	72.5/16.5
PPMI	Normal Control	98	67	31	62/10
	Parkinson	48	35	13	64/11
PSCI	Normal Control	28	16	12	76/6.1
	Stroke	20	15	5	77/5.3

4. Experiments

4.1. Materials

Data Acquisition. Our model is evaluated on three datasets, including two publicly available cohorts and one private clinical collection. Each dataset contains paired resting-state fMRI and DTI scans, detailed as follows. (1) The Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset includes 166 subjects, comprising 115 normal controls (NCs) and 51 individuals diagnosed with mild cognitive impairment (MCI). (2) The Parkinson’s Progression Markers Initiative (PPMI) dataset consists of 146 subjects, including 98 NCs and 48 patients with Parkinson’s disease (PD). (3) The post-stroke cognitive impairment (PSCI) dataset was constructed in-house at the Department of Neuroradiology, Inner Mongolia People’s Hospital, and includes 48 subjects, consisting of 28 NCs and 20 clinically diagnosed stroke disease (SD). All imaging data were parcellated using the AAL atlas, which delineates 116 ROIs: 90 cerebral and 26 cerebellar regions. For the ADNI and PSCI datasets, only the 90 cerebral ROIs were retained to reduce potential noise from cerebellar signals. All 116 ROIs were preserved for PPMI, as cerebellar involvement may contribute to Parkinsonian motor and non-motor manifestations. Comprehensive demographic and clinical details, including gender distribution and deviation age (mean/standard), are summarized in Table 1.

Data Preprocessing. All rs-fMRI data were preprocessed via the Configurable Pipeline for the Analysis of Connectomes (C-PAC) pipeline (Craddock et al., 2013), which involved the removal of the first 10 volumes, slice-timing and motion correction, spatial normalization, nuisance regression with motion parameters, band-pass filtering (0.01–0.1 Hz), and extraction of ROI-level BOLD time series in accordance with the AAL atlas. The DTI data were preprocessed using the PANDA toolbox (Cui et al., 2013), including skull stripping, motion and eddy current correction, registration to T1-weighted MRI, and deterministic fiber tracking with angular (45°) and FA thresholds. Then, structural connectivity matrices were constructed by calculating the mean FA values of streamlines connecting each pair of AAL ROIs.

4.2. Competitive methods

To evaluate the effectiveness of our method, it is compared with three categories of representative baselines: general-purpose GNNs, brain network learning models, and high-order modeling methods. The first category includes GCN (Kipf & Welling, 2017), GAT (Velickovic et al., 2018), GCNII (Chen et al., 2020), GCNH (Cavallo et al., 2023), and CAGNN (Chen et al., 2023). Although initially designed for general graph learning tasks, these models have become essential components of various brain network analysis pipelines owing to their strong expressive power. The second category consists of models explicitly tailored for brain network analysis, including BGAN (Zhou et al., 2024), XG-GNN (Qiu et al., 2024), Cross-GNN (Yang et al., 2024), and OT-MCSTGCN (Zhu et al., 2024), as well as hemispheric-aware methods such as LG-DBGL (Ye et al., 2025) and HeBrainGNN (Shi et al., 2025). The third category focuses on modeling high-order interactions in brain networks, including MHSA (Yuan et al., 2024), CAGT (Pei et al., 2025),

HGST (Han et al., 2026), ALTER (Yu et al., 2024), LHDFormer (Xue et al., 2025b), and DHGFormer (Xue et al., 2025b). These methods leverage hypergraph structures or attention mechanisms to capture complex interactions beyond local neighborhoods.

4.3. Experimental setup

In the experiments, the model was trained using the Adam optimizer with the following common hyperparameter settings: learning rate of $1e-3$, dropout rate of 0.6, weight decay of $1e-4$, 64 hidden units, and 1000 epochs with patience 50. For specific parameters within the MHGCN module, the diffusion time t was selected through a grid search over $[0.5, 10]$ with step size 0.5, the heat-kernel matrix below 0.0001 in $e^{-t\tilde{L}}$ was truncated, the residual propagation coefficients were fixed as $\alpha = 0.1$ and $\beta = 0.5$, and the cross-hemispheric encoder contained two GIN layers with self-loop coefficient initialized to $\epsilon = 0.1$. For the MSDH module, $\gamma = 0.5$ and $\eta = 0.8$ were set. All experiments were carried out using five-fold cross-validation, and the performance was reported in terms of mean Accuracy (Acc), F1-score (F1), and Area Under the Curve (AUC). All baseline methods were implemented according to the hyperparameter configurations reported in the corresponding literature. The entire experimental pipeline was developed in PyTorch and executed on servers equipped with an NVIDIA GeForce RTX 4090 GPU.

5. Results and discussion

5.1. Comparison with competitive methods

Table 2 shows the classification results of LMHGL against competitive methods across three datasets. Compared with the strongest baseline, LMHGL achieves accuracy improvements of 1.65% on ADNI, 2.38% on PPMI, and 2% on PSCI, respectively. In addition, paired t -tests confirm that the accuracy gains are statistically significant with at least $p < .05$. These results demonstrate the effectiveness of our proposed method in brain disorder classification. The reason our method can perform better is that it provides a unified scheme to model lateralized heterogeneity, which captures high-order intra-hemispheric interactions and enables collaboration across views at the decision level. Compared with existing hemispheric-aware methods (e.g., LG-DBGL and HeBrainGNN), which typically extract hemispheric representations into a single embedding, LMHGL preserves hemisphere-specific information and explicitly leverages it with a global representation for collaborative decision prediction, thereby avoiding the loss of discriminative lateralized patterns. Meanwhile, compared with high-order modeling approaches based on hypergraphs (i.e., MHSA and HGST) or attention (i.e., CAGT, ALTER, LHDFormer, and DHGFormer), which rely on data-driven constructions or unconstrained attention, the proposed heat diffusion process models multi-hop dependencies along anatomically plausible structural pathways, leading to more reliable and biologically grounded representations. Therefore, by combining anatomically grounded high-order modeling with explicit lateralization-aware multi-view collaboration decision, LMHGL achieves more discriminative and robust performance for brain disorder classification.

5.2. Ablation study

To assess the contribution of each component in LMHGL, we conduct ablation studies by removing or modifying key modules, as shown in Table 3. First, replacing heat diffusion with the generalized Laplacian (w/o Diffusion) consistently degrades performance, confirming the necessity of modeling high-order interactions beyond first-order message passing. Second, to examine hierarchical representation learning, we remove the inter-layer fusion and use only the final-layer representation $H^{(L)}$ (w/o layer-Fusion). The performance degradation indicates that aggregating varying propagation depths is beneficial for classification. Third, replacing the proposed multi-view attention fusion (w/o

Table 2

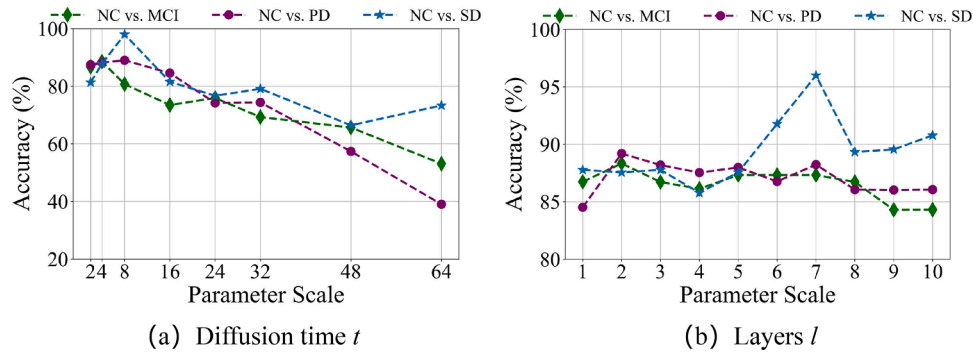
Classification performance comparison (mean/std) across all baselines and datasets. The bold font indicates the best performance term. Statistical significance is indicated by * ($p < .05$) and ** ($p < .01$) based on paired t -tests against our LMHGL model.

Group	Model	ADNI (NC vs. MCI)			PPMI (NC vs. PD)			PSCI (NC vs. SD)		
		ACC (%)	F1 (%)	AUC (%)	ACC (%)	F1 (%)	AUC (%)	ACC (%)	F1 (%)	AUC (%)
General-Purpose GNNs	GCN	74.81/10.8**	73.78/11.1	68.43/12.26	69.87/5.83**	58.78/10.15	54.44/8.89	74.89/5.52**	73.56/4.79	71.83/3.89
	GAT	72.3/2.1**	65.71/5.4	57.67/5.04	69.1/2.02**	59.09/4.86	53.89/3.33	70.67/10.65**	67.3/14.37	68.5/13.42
	GCNII	65.29/9.74**	60.81/6.0	53.7/5.26	71.435/8.14**	61.68/11.47	53.12/3.85	80.0/10.95**	76.75/16.52	77.33/14.29
	GCNH	72.94/6.06**	72.81/5.4	69.58/5.75	73.49/8.97**	83.05/4.61	62.89/15.98	81.11/8.37**	81.13/8.4	83.17/8.54
	CAGNN	70.5/1.97**	60.81/6.0	53.7/5.26	69.87/3.46**	80.53/2.5	57.89/5.46	84.10/4.80**	83.73/5.22	83.89/5.11
Brain Network Learning	BGAN	86.14/5.32*	86.36/5.22	85.65/6.30	77.33/8.58**	75.68/9.95	76.67/8.76	79.12/6.45**	77.52/6.98	78.34/7.25
	XG-GNN	73.57/9.05**	70.49/7.97	60.51/9.46	74.11/15.79**	66.53/22.44	65.95/18.34	71.43/8.01**	69.32/9.25	70.21/8.88
	OT-MCSTGCN	80.71/3.14**	60.57/9.17	74.30/11.31	70.52/4.14**	70.68/11.08	64.95/9.08	81.33/7.42**	79.83/5.35	79.17/11.79
	Cross-GNN	84.95/1.75**	72.47/2.59	79.32/2.26	86.56/10.19*	85.33/8.79	85.12/10.77	87.56/3.81*	83.17/8.38	86.33/5.88
	LG-DBGL	80.18/6.81**	80.25/6.54	80.24/6.4	85.34/6.01*	83.55/7.73	78.33/9.69	83.11/4.55**	83.05/4.89	83.50/5.57
	HeBrainGNN	85.56/4.77**	83.72/5.80	76.55/7.66	83.78/6.93*	82.28/8.57	85.39/6.94	83.11/8.50**	78.67/12.20	82.0/9.27
High-order Brain Network	MHSA	80.75/3.9**	56.33/10.20	69.78/5.39	86.00/6.67*	85.85/6.81	86.69/6.39	91.78/7.56**	91.56/6.62	93.00/6.36
	CAGT	71.43/22.82**	68.89/29.05	86.71/5.20	79.34/19.79**	75.50/25.06	95.75/6.98	96.0/8.01*	95.81/8.29	96.67/10.0
	HGST	86.89/5.16*	83.54/7.25	75.36/8.82	86.42/5.21*	82.37/8.53	83.21/8.33	91.67/4.17*	89.73/5.14	91.89/5.44
	ALTER	81.28/11.28**	79.16/14.38	83.18/6.94	84.58/10.80**	81.29/15.47	87.35/3.43	90.89/7.32*	85.68/12.0	91.67/11.49
	LHDFormer	83.52/5.78**	83.45/9.95	82.87/4.75	84.78/8.26**	83.35/9.94	81.58/13.19	78.76/12.73**	76.85/18.41	88.02/1.67
	DHGFormer	81.21/9.21**	73.15/13.96	85.12/12.22	85.34/6.01*	83.55/7.73	78.33/9.69	64.0/22.89**	56.27/28.29	84.17/17.66
Ours	LMHGL	88.54/4.48	87.42/5.53	81.84/8.22	88.94/2.86	89.04/3.13	88.95/5.3	98.0/4.0	97.93/4.13	97.5/5.0

Table 3

The ablation results (mean/std) of our proposed method under different variants.

Model	ADNI (NC vs. MCI)			PPMI (NC vs. PD)			PSCI (NC vs. SD)		
	ACC (%)	F1 (%)	AUC (%)	ACC (%)	F1 (%)	AUC (%)	ACC (%)	F1 (%)	AUC (%)
w/o Diffusion	87.34/4.03	86.0/4.91	79.36/6.73	79.26/12.74	78.88/13.04	75.56/14.84	87.78/7.44	87.27/7.83	85.53/8.58
w/o Layer-Fusion	85.56/5.81	86.93/3.67	86.38/4.60	86.75/1.92	86.81/1.94	86.70/3.68	88.0/11.66	87.44/12.24	86.67/13.54
w/o View-Fusion	86.35/4.98	85.12/5.01	82.38/4.21	84.76/2.18	85.44/3.01	85.06/2.13	82.04/7.39	84.38/7.25	84.65/5.44
w/o MVDM	75.94/8.23	74.48/7.45	69.23/8.1	70.50/12.73	70.41/12.2	69.59/12.98	85.56/4.55	85.08/4.99	84.17/6.12
LMHGL	88.54/4.48	87.42/5.53	81.84/8.22	88.94/2.86	89.04/3.13	88.95/5.3	98.0/4.0	97.93/4.13	97.5/5.0

**Fig. 3.** Sensitivity analysis of diffusion time and the number of convolution layers.

View-Fusion) with additive attention (i.e., $\alpha_k = \text{softmax}(q^T \tanh(WZ_k))$) results in a performance decline, suggesting that the lightweight fusion strategy provides more stable view weighting under limited data. Finally, removing MVDM (w/o MVDM) and relying solely on the global embedding Z_G , which causes the most significant degradation, demonstrates that directly relying on a unified global representation is insufficient, while explicitly incorporating hemispheric-specific representations at the decision level is critical for effective prediction. These results validate each component's effectiveness and support the design rationale of LMHGL in modeling lateralized brain networks.

5.3. Parameter sensitivity analysis

This study investigates the sensitivity of LMHGL concerning four key hyperparameters that may affect model performance: the diffusion time t , the number of convolutional layers l in the Intra-Inter GCN, and the scaling factors γ and η . The sensitivity analyses are performed under the

same five-fold cross-validation protocol. As demonstrated in Fig. 3(a), the model achieves optimal performance when $t \in [2, 5]$, suggesting that moderate diffusion effectively balances local structural fidelity and high-order information propagation. Lower values of t constrain the diffusion range and limit high-order interaction modeling, whereas excessively high values of t result in over-smoothing effects that reduce discriminative capability. Fig. 3(b) indicates that increasing network depth initially enhances performance by allowing for richer hierarchical feature extraction within hemispheric subnetworks, but excessive depth results in over-smoothing and stability degradation. Fig. 4 shows a joint sensitivity analysis of the decorrelation scaling factors γ and η . The results indicate that LMHGL achieves stable and relatively strong performance when η is set to a greater value and γ remains moderate. This implies that maintaining sufficient global contextual information while partially diminishing redundant global components in hemispheric representations is advantageous for collaborative decision-making. Overall, these analyses offer practical guidance for parameter tuning and enhance our understanding of the model's behavior.

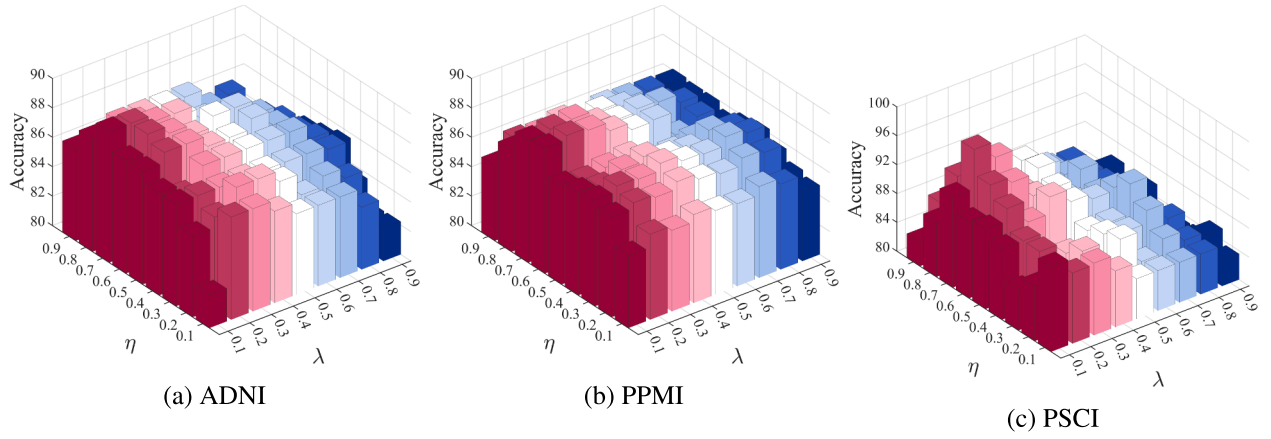


Fig. 4. Sensitivity analysis of scaling factor γ and η across all datasets.

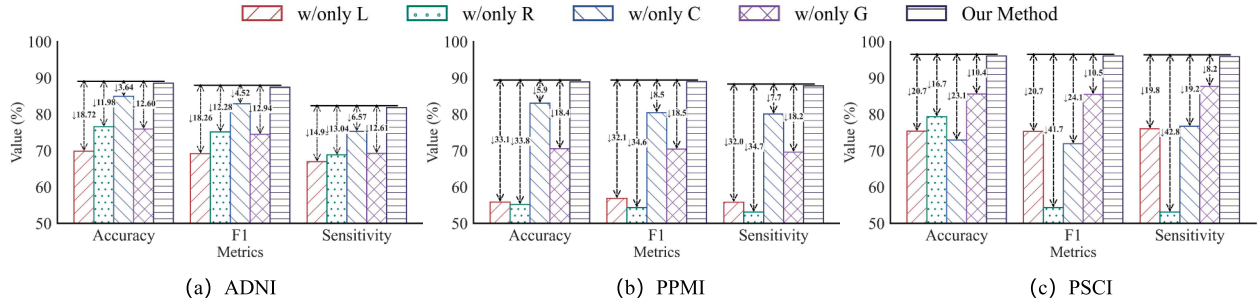


Fig. 5. The lateralization-guided ablation study of our proposed method under different subnetwork views.

5.4. Lateralized heterogeneity validation

To investigate the role of lateralized modeling, we conduct a hemisphere-guided ablation study by retaining only a single view, including the left (w/only L), right (w/only R), cross-hemispheric (w/only C), and fused global (w/only G) representation. Since only one view is used in each setting, the MVDM module is disabled. As shown in Fig. 5, all single-view variants (w/only L, w/only R, and w/only C) underperform the global baseline (w/only G), indicating that disease-related patterns are distributed across multiple regions and cannot be fully captured by any single hemispheric view. However, the w/only G performs worse than the LMHGL model, suggesting that directly aggregating whole-brain information is insufficient and may obscure lateralized pathological signals across hemispheres. By jointly modeling left, right, and cross-hemispheric subnetworks and integrating them through MVDM module, the LMHGL captures complementary information from different views while preserving hemispheric distinctions. This enables the model to adaptively leverage asymmetric patterns when present, rather than suppressing them. Overall, these findings provide empirical evidence that modeling hemispheric heterogeneity is essential for accurate brain network analysis.

5.5. High-order interactions evaluation

In LMHGL, high-order interactions are modeled via anatomically constrained heat diffusion along the structural connectome. To validate its effectiveness, we conduct two complementary analyses from functional alignment and spectral perspectives.

Functional Alignment Validation. To check whether the diffused graph reflects biologically meaningful interactions, we compare three types of matrices for each hemisphere: the structural adjacency matrix A^s , the diffused matrix $P = e^{-t\hat{L}}$, and the empirical functional connectivity matrix A^f derived from resting-state fMRI. For each pair of matrices, the Pearson correlation coefficient (PCC) and Kullback-Leibler

Table 4

Functional alignment analysis between structural, diffused, and functional connectivity matrices using PCC and KL divergence. (L/R: Left/Right hemisphere).

Metric	Group	ADNI		PPMI		PSCI	
		L	R	L	R	L	R
PCC	A^s vs. A^f	0.629	0.630	0.644	0.654	0.649	0.653
	P vs. A^f	0.699	0.700	0.727	0.730	0.733	0.737
	P vs. A^s	0.936	0.940	0.919	0.929	0.934	0.936
KL	A^s vs. A^f	0.025	0.023	0.025	0.025	0.023	0.022
	P vs. A^f	0.004	0.005	0.006	0.007	0.005	0.004
	P vs. A^s	0.022	0.020	0.027	0.027	0.020	0.021

(KL) divergence are computed. As shown in Table 4, P remains highly consistent with A^s , indicating that diffusion preserves the underlying anatomical structure. At the same time, P exhibits stronger similarity to A^f than A^s , suggesting that diffusion transforms structural connectivity into a representation that better reflects functional coupling patterns. This result demonstrates that heat diffusion captures high-order interactions that bridge anatomical structure and functional dynamics, while remaining constrained by biologically plausible pathways.

Spectral Reorganization Analysis. We further examine how diffusion reshapes topological structure to characterize high-order interactions through spectral analysis. As reported in Table 5, P shows a significant reduction in spectral entropy ($p < .05$), indicating a transition from locally irregular structures to more globally organized connectivity. In addition, the eigenvalues of P shift toward lower frequencies, reflecting the attenuation of high-frequency (local) variations and the dominance of low-frequency (global) modes. These spectral changes suggest that diffusion suppresses isolated local interactions while enhancing coherent global dependencies, thereby facilitating long-range, multi-hop information propagation. This provides further evidence that the pro-

Table 5

Spectral reorganization analysis of structural and diffused connectivity matrices in terms of spectral entropy and eigenvalues. (L/R: Left/Right-hemispheres; * indicates significant differences vs. A^s , paired t-test, $p < .05$).

Metric	Group	ADNI		PPMI		PSCI	
		L	R	L	R	L	R
Spectral Entropy	A^s	3.265	3.235	3.329	3.340	3.282	3.272
	P	2.393*	2.388*	2.505*	2.570*	2.379*	2.401*
Spectral Eigenvalue	A^s	0.695	0.676	0.698	0.674	0.704	0.692
	P	0.312*	0.330*	0.318*	0.337*	0.307*	0.318*

posed diffusion process effectively captures high-order interactions in brain networks.

Overall, these results provide empirical evidence that anatomically constrained heat diffusion effectively models high-order interactions in brain networks while maintaining biological plausibility.

5.6. Disease-related ROIs identification

In this subsection, disease-related ROIs are analyzed by quantifying each subject's average node-level topological changes score, specifically degree changes, by comparing the heat-diffused matrix in Eq. (5) with its structural connectivity matrix A^s . This score indicates the extent of alteration in a node's connectivity profile when high-order interactions are integrated, thereby highlighting ROIs that demonstrate significant topological reorganization. The top 10 ROIs with the highest scores for each hemisphere are listed in Table 6, with representative disease-relevant ROI underlined and their corresponding spatial distributions are visualized in Fig. 6.

Clinical Relevance. The identified key ROIs align well with established neuroimaging findings. In the ADNI dataset, the postcentral gyrus (PCG) has been implicated in somatosensory deficits in early AD (Wang et al., 2006); the superior frontal gyrus medial (SFGmed) is part of the default mode network (DMN) and shows early amyloid deposition (Palmqvist et al., 2017); the paracentral lobule (PCL) and cuneus (CUN) are involved in visuospatial and motor dysfunction, respectively. For the PPMI dataset, the Vermis regions (Verm6-10) and cerebellar Crus regions (Cere7b, Cere9) are critical for motor coordination and have been consistently reported to show structural and functional alterations in PD (Wu & Hallett, 2013). The involvement of SFGmed in PD is associated with non-motor symptoms such as executive dysfunction (Kehagia et al., 2013). For post-stroke cognitive impairment (PSCI dataset), the PCG, CUN, PCL, and anterior cingulate gyrus (ACG) are regions frequently affected by stroke lesions and are linked to sensorimotor and attentional deficits (Mijajlović et al., 2017; Yue et al., 2023). The pallidum (PAL) has been implicated in post-stroke recovery mechanisms (Grefkes & Fink, 2011). Notably, our model also identifies regions that

Table 6

Top-10 ROIs identified by average diffusion-induced degree change across subjects.

	ADNI		PPMI		PSCI	
	ROIs	Score	ROIs	Score	ROIs	Score
R	<u>PCG.R</u>	0.4765	<u>Verm6</u>	0.5763	<u>PCG.R</u>	0.5033
	<u>ORBmid.R</u>	0.4743	<u>Verm8</u>	0.5447	<u>CUN.R</u>	0.4741
	<u>SFGmed.R</u>	0.4718	<u>Verm3</u>	0.5371	<u>SFGmed.R</u>	0.4678
	<u>PCL.R</u>	0.4539	<u>Verm10</u>	0.5310	<u>ORBmid.R</u>	0.4614
	<u>HES.R</u>	0.4479	<u>ORBmid.R</u>	0.4946	<u>PCL.R</u>	0.4505
	<u>SMA.R</u>	0.4420	<u>PCG.R</u>	0.4946	<u>HES.R</u>	0.4431
	<u>OLF.R</u>	0.4178	<u>SFGmed.R</u>	0.4776	<u>REC.R</u>	0.4355
	<u>CUN.R</u>	0.4161	<u>Cere9.R</u>	0.4657	<u>SMA.R</u>	0.4291
	<u>REC.R</u>	0.4000	<u>Cere7b.R</u>	0.4608	<u>OLF.R</u>	0.4246
	<u>ACG.R</u>	0.3972	<u>HES.R</u>	0.4553	<u>PAL.R</u>	0.4174
L	<u>PCG.L</u>	0.4724	<u>Verm9</u>	0.5904	<u>PCG.L</u>	0.5096
	<u>PCL.L</u>	0.4715	<u>Verm7</u>	0.5783	<u>CUN.L</u>	0.4830
	<u>ORBmid.L</u>	0.4568	<u>Verm1_2</u>	0.5752	<u>PCL.L</u>	0.4555
	<u>CUN.L</u>	0.4259	<u>Verm4_5</u>	0.5568	<u>ORBmid.L</u>	0.4481
	<u>HES.L</u>	0.4222	<u>PCG.L</u>	0.4571	<u>HES.L</u>	0.4313
	<u>SFGmed.L</u>	0.4190	<u>ORBmid.L</u>	0.4496	<u>REC.L</u>	0.4291
	<u>REC.L</u>	0.4115	<u>CUN.L</u>	0.4452	<u>OLF.L</u>	0.4165
	<u>SMA.L</u>	0.4063	<u>PCL.L</u>	0.4435	<u>PAL.L</u>	0.4135
	<u>OLF.L</u>	0.4061	<u>HES.L</u>	0.4363	<u>ACG.L</u>	0.4127
	<u>PAL.L</u>	0.3998	<u>ACG.L</u>	0.4326	<u>SFGmed.L</u>	0.4097

are less commonly reported (e.g., OLF, REC in ADNI; Cere9 in PPMI), which may represent novel biomarkers worthy of further investigation. **Statistical validation.** To validate the statistical significance of these disease-related ROIs, we performed independent two-sample t-tests comparing the diffusion-induced degree change between patient and healthy control groups. As shown in Table 7, most ROIs exhibited significant differences in degree change between patients and controls ($p < 0.05$). This confirms that the topological reorganization captured by our heat diffusion module is not random but systematically differs between diagnostic groups.

Overall, these quantitative results confirm that the identified ROIs are indeed disease-relevant and that LMHGL effectively captures meaningful topological reorganization.

5.7. Evaluation of model robustness

To evaluate the robustness of LMHGL, we perform edge removal experiments on all three datasets by progressively discarding the weakest connections (i.e., lowest FA values) from the structural adjacency matrix A^s at ratios of 5%, 10%, 15%, 20%, and 25%. The model is then re-evaluated under identical settings. As shown in Fig. 7, after removing 5% of the weakest edges, LMHGL exhibited only a slight drop in accuracy across all three datasets. Even after removing 10% or 15% of edges, the model still maintained a high level of performance, demonstrating notable resilience to structural connectivity incompleteness. When the

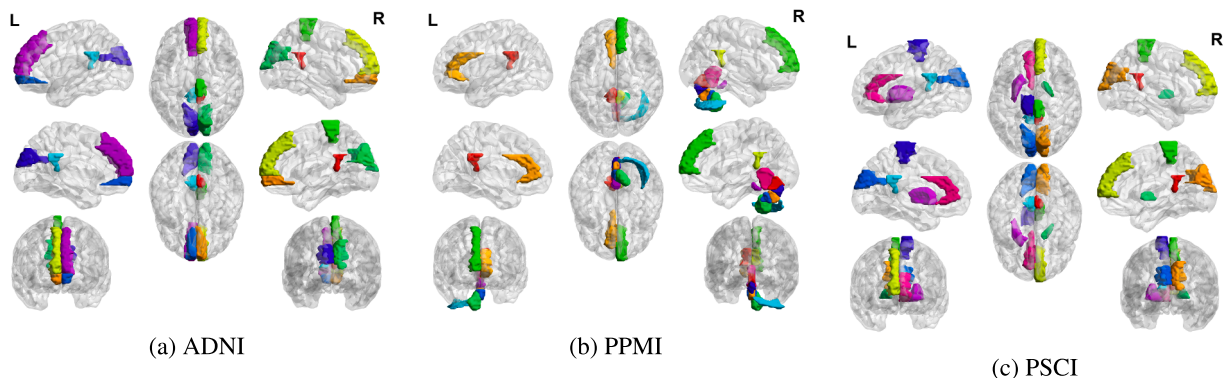


Fig. 6. Anatomical locations of disease-relevant ROIs identified in the ADNI, PPMI, and PSCI datasets.

Table 7

Statistical validation of disease-related ROIs between patients and healthy controls across three datasets. Statistical significance is indicated by * ($p < .05$), ** ($p < .01$), and - denotes not significant based on independent two-sample t-tests. The sign of t-statistics reflects the direction of group differences.

ADNI			PPMI			PSCI		
ROI	t-statistic	p-value	ROI	t-statistic	p-value	ROI	t-statistic	p-value
PCG.R	3.4063	**	Verm6	2.5240	*	PCG.R	2.8164	**
ORBmid.R	-3.1594	**	Verm8	-0.4588	-	CUN.R	-0.1197	-
SFGmed.R	-3.7462	**	PCG.R	2.0875	*	SFGmed.R	-2.4337	*
PCL.R	1.3591	-	SFGmed.R	2.2362	*	PCL.R	0.9616	-
CUN.R	0.3669	-	Cere9.R	1.5711	*	PAL.R	3.2102	**
PCG.L	2.6324	**	Cere7b.R	1.8887	*	PCG.L	2.1217	*
PCL.L	1.1354	-	Verm9	-0.2515	-	CUN.L	-0.2276	-
ORBmid.L	-1.4362	*	Verm7	3.7147	**	PCL.L	0.6053	-
CUN.L	1.5720	*	Verm1_2	1.9914	*	PAL.L	4.5284	**
SFGmed.L	-2.0254	*	Verm4_5	1.9342	*	ACG.L	2.0666	*

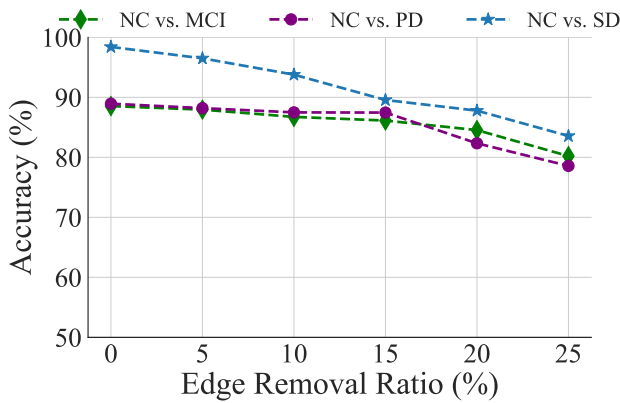


Fig. 7. Robustness of LMHGL to progressive edge removal on the ADNI, PPMI, and PSCI datasets.

removal ratio increased to 20% and 25%, a more pronounced performance decline was observed, which is expected as excessive removal inevitably destroys critical anatomical pathways. Overall, these results demonstrate that LMHGL is robust to moderate levels of structural edge removal, which is particularly valuable for clinical applications where DTI-derived connectivity may be incomplete or noisy.

5.8. Time complexity analysis

We analyzed the time complexity of the proposed LMHGL framework. For a brain network with N nodes, E edges, and feature dimension d , the TNDS first partitions the network into G_L , G_R , and G_C through simple indexing and slicing. Since this preprocessing step is performed only once before training, its computational cost is negligible. In the MHGCN module, for the hemispheric subnetworks G_L and G_R , the diffusion process in Eq. (5) has a time complexity of $O(k(|E_L| +$

$|E_R|)d)$. The L -layer intra-inter GCN defined in Eqs. (6)-(7) requires $O(L(|E_L|d + N_L d^2)) + O(L(|E_R|d + N_R d^2))$. The cross-hemispheric encoder has a complexity of $O(L(|E_C|d + N_C d^2))$. The remaining computations in MHGCN and all operations within MVDM are mostly lightweight MLP and Softmax transformations, each with $O(d^2)$, which does not affect the overall asymptotic cost. Combining all terms, the total time complexity of our framework can be expressed as $O(k(|E_L| + |E_R|)d + l(|E_L|d + N_L d^2 + |E_R|d + N_R d^2 + |E_C|d + N_C d^2))$. Since $|E_L| \approx |E_R| \approx |E|/2$ and $|E_C| \ll |E|$ in typical brain networks, the overall complexity is dominated by the graph propagation and can be simplified as $O(l|E|d + lN d^2)$, indicating the model's scalability for brain disease classification. Furthermore, Table 8 reports the model complexity of LMHGL on a per-subject basis, in terms of parameter count (M), inference time (ms), and GPU memory usage (MB). The results show that LMHGL achieves substantially lower parameter complexity, maintaining competitive inference time and moderate memory consumption.

6. Limitation and future work

6.1. Limitation to large-scale validation

We note that the datasets used in this study are relatively small, particularly the private PSCI dataset, which may limit the statistical power and generalizability of the results. To further evaluate LMHGL on a larger cohort, we conducted additional experiments on the publicly available ABIDE dataset (Xu et al., 2023), which contains approximately 1025 subjects. However, ABIDE provides only resting-state fMRI without matched DTI data. Since LMHGL relies on structural connectivity for its anatomically constrained heat diffusion, we used functional connectivity (Pearson correlation) as a surrogate. As shown in Table 9, LMHGL achieved an accuracy of 69.49% under this setting. This performance drop is expected because the anatomical constraint is lost, and the diffusion process degenerates into unconstrained smoothing. Therefore, this experiment does not reflect the model's true scalability but rather its sensitivity to the availability of DTI data. A proper large-scale

Table 8

Model complexity comparison on a per-subject basis across different datasets.

Method	ADNI (NC vs. MCI)			PPMI (NC vs. PD)			PSCI (NC vs. SD)		
	Params (M)	Inference (ms)	GPU Mem (MB)	Params (M)	Inference (ms)	GPU Mem (MB)	Params (M)	Inference (ms)	GPU Mem (MB)
LG-DBGL	0.0055	1.53	37.63	0.0068	1.54	44.04	0.0055	1.52	25.16
HeBrainGNN	0.1061	121.18	63.27	0.1253	169.92	77.65	0.0752	120.09	29.19
CAGT	1.8639	25.87	48.10	2.0836	24.87	78.11	1.8639	24.33	51.21
HGST	2.9965	4.14	71.24	2.9965	4.52	89.82	2.9965	3.86	83.35
ALTER	3.2009	8.45	79.95	3.2571	10.96	81.18	3.2009	8.29	79.95
LHDFormer	1.5286	17.44	36.50	2.1108	16.16	45.24	1.5286	16.86	36.56
DHGFormer	0.2341	6.41	21.68	0.3078	6.67	23.33	0.2341	6.55	21.68
LMHGL	0.0158	4.70	63.51	0.0179	4.66	79.96	0.0125	4.76	27.88

Table 9
Classification results (mean/std) on ABIDE dataset.

Model	ABIDE		
	ACC (%)	F1 (%)	AUC (%)
LG-DBGL	63.33/4.47	61.36/5.98	69.54/4.18
HeBrainGNN	64.0/2.36	63.54/2.54	63.81/2.52
MHSA	66.56/3.13	68.83/2.53	68.68/2.03
CAGT	71.85/2.91	71.61/3.0	73.20/2.79
HGST	67.99/3.03	67.38/3.28	67.88/3.18
ALTER	71.15/2.31	71.41/3.02	71.06/1.54
LHDFormer	72.32/1.08	72.21/1.40	74.84/1.51
DHGFFormer	71.53/2.58	71.52/2.45	76.18/4.03
LMHGL	69.49/2.14	69.21/2.09	69.48/1.95

validation requires datasets containing both fMRI and DTI modalities, such as the UK Biobank¹ dataset with more than 10,000 participants. We plan to pursue this direction in future work.

6.2. Future directions toward pretraining

Although LMHGL achieves strong performance, it is primarily designed for supervised learning. Recent advances in graph pretraining have shown that self-supervised strategies can learn generalizable and transferable representations from large-scale neuroimaging data. For example, BrainGFM (Wei et al., 2026) integrates graph contrastive learning with masked autoencoding for large-scale fMRI pretraining, while MCSP (Wei et al., 2025) develops a multi-modal self-supervised framework that jointly models spatial, temporal, and spectral information in fMRI and EEG fusion tasks. In contrast, LMHGL is built upon biologically grounded inductive biases, particularly hemispheric-aware modeling and anatomically constrained heat diffusion. Therefore, extending LMHGL into a self-supervised setting offers a promising direction to introduce neuroanatomical priors and improve sensitivity to lateralized pathological patterns.

From this perspective, LMHGL can be extended toward self-supervised learning in two directions. First, the heat diffusion mechanism enables structure-aware pretraining objectives by defining multi-hop consistency over the structural connectome. Specifically, diffusion can generate multi-scale views of the same graph under progressive propagation, allowing the model to learn stable representations along anatomically plausible pathways. This provides a principled way to encode long-range dependencies during pretraining. Second, the hemispheric decomposition offers a biologically grounded multi-view setting for contrastive learning. By treating the left, right, and cross-hemispheric subnetworks as intrinsic views of the same subject, self-supervised objectives can be formulated to align these representations while preserving their differences. This enables the model to capture both shared global structures and lateralized variations for brain disorders. Overall, these directions provide feasible pathways for extending LMHGL toward a pretraining paradigm. Future work will focus on integrating anatomically constrained modeling with self-supervised learning to further improve generalization and clinical applicability.

7. Conclusion

In this paper, we propose LMHGL for brain disorder classification, which provides a unified scheme to model hemispheric asymmetry. Unlike existing lateralization-aware methods that rely on standard message passing within hemispheres, LMHGL introduces anatomically constrained heat diffusion to explicitly model high-order synergistic interactions along structural connectivity. Furthermore, a multi-view decision mechanism is designed to preserve hemisphere-specific pathological patterns and adaptively integrate predictions from left, right, and

global views to collaboratively make decisions. Extensive experiments on three real-world datasets demonstrate that LMHGL consistently outperforms a wide range of baselines. Ablation studies validate the contribution of each key component, while functional alignment and spectral analyses confirm the biological plausibility of the proposed heat diffusion. Moreover, the identified disease-related ROIs align well with clinical literature, and robustness experiments show that LMHGL tolerates moderate incompleteness of structural connectivity. These results collectively highlight that explicitly modeling hemispheric asymmetry and anatomically constrained high-order interactions provides a principled and effective strategy for discriminative brain network representation. Future work will extend LMHGL to self-supervised pretraining paradigms and evaluate it on larger multi-modal cohorts.

CRedit authorship contribution statement

Jiazhen Ye: Writing – review & editing, Methodology, Software, Formal analysis, Visualization; **Manman Yuan:** Conceptualization, Resources, Supervision, Funding acquisition, Project administration; **Yan Zhao:** Data Curation, Investigation; **Can Yin:** Validation, Investigation; **Mengyi Shao:** Data curation, Formal analysis; **Jürgen Kurths:** Supervision, Project administration.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant 62466042.

References

- Abdelnour, F., Voss, H. U., & Raj, A. (2014). Network diffusion accurately models the relationship between structural and functional brain connectivity networks. *NeuroImage*, 90, 335–347.
- Cavallo, A., Grohnfeldt, C., Russo, M., Lovisotto, G., & Vassio, L. (2023). GCNH: A simple method for representation learning on heterophilous graphs. In *ICNEN* (pp. 1–8).
- Chen, J., Chen, S., Gao, J., Huang, Z., Zhang, J., & Pu, J. (2024). Exploiting neighbor effect: Conv-agnostic GNN framework for graphs with heterophily. *IEEE Transactions on Neural Networks and Learning Systems*, 35, 13383–13396.
- Chen, M., Wei, Z., Huang, Z., Ding, B., & Li, Y. (2020). Simple and deep graph convolutional networks. In *ICML*.
- Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S. S., Yan, C., Li, Q., Lurie, D., Vogelstein, J., Burns, R. et al. (2013). Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (C-PAC). *Frontiers in Neurology*. <https://doi.org/10.3389/conf.fninf.2013.09.00042>.
- Cui, Z., Zhong, S., Xu, P., Gong, G., & He, Y. (2013). Panda: A pipeline toolbox for analyzing brain diffusion images. *Frontiers in Human Neuroscience*, 7–2013. <https://doi.org/10.3389/fnhum.2013.00042>.
- Fang, H., Wang, H., Gao, Y., Zhang, Y., Bu, J., Han, B., & Lin, H. (2025). InsGNN: Interpretable spatio-temporal graph neural networks via information bottleneck. *Information Fusion*, 119, 102997.
- Fernández-Carriba, S., Loeches, Á., Morcillo, A., & Hopkins, W. D. (2002). Asymmetry in facial expression of emotions by chimpanzees. *Neuropsychologia*, 40(9), 1523–1533.
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. In *ICML* (pp. 1263–1272). (vol. 70).
- Grefkes, C., & Fink, G. R. (2011). Reorganization of cerebral networks after stroke: New insights from neuroimaging with connectivity approaches. *Brain*, 134(5), 1264–1276.
- Gu, Y., Peng, S., Li, Y., Gao, L., & Dong, Y. (2024). FC-HGNN: A heterogeneous graph neural network based on brain functional connectivity for mental disorder identification. *Information Fusion*, 113, 102619.
- Han, X., Lei, M., & Li, J. (2026). Hypergraph-based semantic and topological self-supervised learning for brain disease diagnosis. *Pattern Recognition*, 169, 111921.
- Hao, X., Li, J., Ma, M., Qin, J., Zhang, D., Liu, F., Initiative, A. D. N. et al. (2024). Hypergraph convolutional network for longitudinal data analysis in Alzheimer's disease. *Computers in Biology Medicine*, 168, 107765.

¹ <https://www.ukbiobank.ac.uk/>.

- Hemker, K., Simidjievski, N., & Jamnik, M. (2024). Healnet: Multimodal fusion for heterogeneous biomedical data. In *NeurIPS* (pp. 64479–64498). (vol. 37).
- Jiang, X., Ding, X., Xia, Z., Wang, H., & Jiao, Z. (2025). Self-perceptive feature fusion network with multi-channel graph convolution for brain disorder diagnosis. *Expert Systems with Applications*, 284, 127984.
- Kan, X., Dai, W., Cui, H., Zhang, Z., Guo, Y., & Yang, C. (2022). Brain network transformer. In *NeurIPS*.
- Kehagia, A. A., Barker, R. A., & Robbins, T. W. (2013). Cognitive impairment in parkinson's disease: The dual syndrome hypothesis. *Neurodegenerative Diseases*, 11(2), 79–92.
- Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *ICLR*.
- Kondor, R. (2002). Diffusion kernels on graphs and other discrete structures. In *ICML*.
- Lee, D.-J., Shin, D.-H., Son, Y.-H., Han, J.-W., Oh, J.-H., Kim, D.-H., Jeong, J.-H., & Kam, T.-E. (2024). Spectral graph neural network-based multi-atlas brain network fusion for major depressive disorder diagnosis. *IEEE Journal of Biomedical and Health Informatics*, 28, 2967–2978.
- Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L. H., Ventola, P., & Duncan, J. S. (2021). BrainGNN: Interpretable brain graph neural network for fMRI analysis. *Medical Image Analysis*, 74, 102233.
- Liu, J., Cui, W., Chen, Y., Ma, Y., Dong, Q., Cai, R., Li, Y., & Hu, B. (2024). Deep fusion of multi-template using spatio-temporal weighted multi-hypergraph convolutional networks for brain disease analysis. *IEEE Transactions on Medical Imaging*, 43(2), 860–873.
- Luo, X., Wu, J., Yang, J., Xue, S., Beheshti, A., Sheng, Q. Z., McAlpine, D., Sowman, P. F., Giral, A., & Yu, P. S. (2024). Graph neural networks for brain graph learning: A survey. In *IJCAI* (pp. 8170–8178).
- Mijajlović, M. D., Pavlović, A., Brainin, M., Heiss, W.-D., Quinn, T. J., Ihle-Hansen, H. B., Hermann, D. M., Assayag, E. B., Richard, E., Thiel, A. et al. (2017). Post-stroke dementia—a comprehensive review. *BMC Medicine*, 15(1), 11.
- Palmqvist, S., Schöll, M., Strandberg, O., Mattsson, N., Stomrud, E., Zetterberg, H., Blennow, K., Landau, S., Jagust, W., & Hansson, O. (2017). Earliest accumulation of β -amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nature Communications*, 8(1), 1214.
- Pei, S., Ma, J., Lv, Z., Zhang, C., & Guan, J. (2025). Community-aware graph transformer for brain disorder identification. In *IJCAI* (pp. 4191–4199).
- Qiu, X., Wang, F., Sun, Y., Lian, C., & Ma, J. (2024). Towards graph neural networks with domain-generalizable explainability for fMRI-based brain disorder diagnosis. In *MICCAI* (pp. 454–464).
- Shi, G., Zhu, Y., Liu, W., Yao, Q., & Li, X. (2025). Heterogeneous graph-based multimodal brain network learning. *IEEE Transactions on Knowledge and Data Engineering*, 37(8), 4664–4676.
- Sim, J., Lee, M., Wu, G., & Kim, W. H. (2024). Multi-modal graph neural network with transformer-guided adaptive diffusion for preclinical Alzheimer classification. In *MICCAI*. (vol. LNCS 15005).
- Sun, Y., Chen, Y., Collinson, S. L., Bezerianos, A., & Sim, K. (2017). Reduced hemispheric asymmetry of brain anatomical networks is linked to schizophrenia: A connectome study. *Cerebral Cortex*, 27(1), 602–615.
- Thapaliya, B., Akbas, E., Chen, J., Sapkota, R., Ray, B., Suresh, P., Calhoun, V. D., & Liu, J. (2025). Brain networks and intelligence: A graph neural network based approach to resting state fMRI data. *Medical Image Analysis*, 101, 103433.
- Thiebaut de Schotten, M., Dell'Acqua, F., Forkel, S., Simmons, A., Vergani, F., Murphy, D. G. M., & Catani, M. (2011). A lateralized brain network for visuo-spatial attention. *Nature Neuroscience*, 14, 1245–1246.
- Toga, A. W., & Thompson, P. M. (2003). Mapping brain asymmetry. *Nature Reviews Neuroscience*, 4, 37–48.
- 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, 273–289.
- Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In *ICLR*.
- Wang, L., Zang, Y., He, Y., Liang, M., Zhang, X., Tian, L., Wu, T., Jiang, T., & Li, K. (2006). Changes in hippocampal connectivity in the early stages of Alzheimer's disease: Evidence from resting state fMRI. *NeuroImage*, 31(2), 496–504.
- Wei, X., Zhao, K., Jiao, Y., Carlisle, N. B., Xie, H., Fonzo, G. A., & Zhang, Y. (2025). Multi-modal cross-domain self-supervised pre-training for fMRI and EEG fusion. *Neural Networks*, 184, 107066.
- Wei, X., Zhao, K., Jiao, Y., He, L., & Zhang, Y. (2026). A brain graph foundation model: Pre-training and prompt-tuning across broad atlases and disorders. In *ICLR*.
- Wen, G., Cao, P., Bao, H., Yang, W., Zheng, T., & Zaiane, O. (2022). MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. *Computers in Biology and Medicine*, 142, 105239.
- Wu, T., & Hallett, M. (2013). The cerebellum in Parkinson's disease. *Brain*, 136(3), 696–709.
- Wu, X., Kong, X., Vatansever, D., Liu, Z., Zhang, K., Sahakian, B., Robbins, T., Feng, J., Thompson, P., & Zhang, J. (2022). Dynamic changes in brain lateralization correlate with human cognitive performance. *PLoS Biology*, 20, 1–28.
- Xu, J., Yang, Y., Huang, D., Gururajapathy, S. S., Ke, Y., Qiao, M., Wang, A., Kumar, H., McGeown, J., & Kwon, E. (2023). Data-driven network neuroscience: On data collection and benchmark. In *NeurIPS* (pp. 21841–21856). (vol. 36).
- Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? In *ICLR*.
- Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K.-i., & Jegelka, S. (2018). Representation learning on graphs with jumping knowledge networks. In *ICML* (pp. 5453–5462).
- Xue, R., Hu, H., Zhang, Z., Han, X., Wang, J., Gao, Y., Du, S., DHGFormer: Dynamic hierarchical graph transformer for disorder brain disease diagnosis, in: *MICCAI*, 2025b, pp. 268–278.
- Xue, R., Han, X., Hu, H., Zhang, Z., Du, S., & Gao, Y. (2025b). Adaptive embedding for long-range high-order dependencies via time-varying transformer on fMRI. In *MICCAI* (pp. 46–55).
- Yang, Y., Ye, C., Guo, X., Wu, T., Xiang, Y., & Ma, T. (2024). Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning. *IEEE Transactions on Medical Imaging*, 43(1), 108–121.
- Ye, J., Yuan, M., Li, J., Jia, W., Wang, J., & Li, J. (2025). LG-DBGL: Lateralization-guided dissociative brain graph learning for Alzheimer's disease identification. In *MICCAI*.
- Yu, S., Jin, S., Li, M., Sarwar, T., & Xia, F. (2024). Long-range brain graph transformer. In *NeurIPS* (pp. 24472–24495). (vol. 37).
- Yuan, M., Jia, W., Luo, X., Ye, J., Zhu, P., & Li, J. (2024). MHSA: A multi-scale hypergraph network for mild cognitive impairment detection via synchronous and attentive fusion. In *BIBM* (pp. 2808–2815).
- Yue, X., Li, Z., Li, Y., Gao, J., Han, H., Zhang, G., Li, X., Shen, Y., Wei, W., Bai, Y. et al. (2023). Altered static and dynamic functional network connectivity in post-stroke cognitive impairment. *Neuroscience Letters*, 799, 137097.
- Zhou, Z., Wang, Q., An, X., Chen, S., Sun, Y., Wang, G., & Yan, G. (2024). A novel graph neural network method for Alzheimer's disease classification. *Computers in Biology and Medicine*, 180, 108869.
- Zhu, Q., Li, S., Meng, X., Xu, Q., Zhang, Z., Shao, W., & Zhang, D. (2024). Spatio-temporal graph hubness propagation model for dynamic brain network classification. *IEEE Transactions on Medical Imaging*, 43(6), 2381–2394.