Information Fusion 135 (2026) 104426

Contents lists available at ScienceDirect

Information Fusion

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

Full Length Article

Learning brain structure-function interactions via modularity-guided attention-based graph neural networks

Jing Xia [a,b], Yi Hao Chan [b], Deepank Girish [b], Conghao Wang [b], Jagath C. Rajapakse [b,*]

[a] Key Laboratory for Biomedical Engineering of Ministry of Education, Zhejiang Key Laboratory of Intelligent Sensing Technology and Advanced Medical Instrument, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China
[b] College of Computing and Data Science, Nanyang Technological University, Singapore

ARTICLE INFO

Keywords:
Structure-function coupling
Functional and structural connectivity
Fluid cognition
Parkinson's disease
Schizophrenia
Modularity
Attention

ABSTRACT

Incorporating brain structure-function interactions has been shown to enhance brain disorder diagnosis and cognitive prediction. However, current studies primarily focus on one-to-one interactions between corresponding regions of interest (ROIs), overlooking more complex interactions across different ROIs. Since ROIs within the same module exhibit more similar neural mechanisms of structure-function coupling compared to those in different modules, modular-level homogeneity provides stronger biological guidance for learning complex interactions. Though leveraging this guidance can further improve prediction performance, such studies remain scarce. To address this gap, we propose a modularity-guided attention-based graph neural network that extracts and emphasizes key structure-function interactions across ROIs relevant to specific tasks. Specifically, a modularity-guided interactive network is designed to define modularity-specific convolution operations learning interactions across structural and functional ROIs based on modular homogeneity. A modular interactive edge attention model is designed to highlight the most salient interactions by using a cross-attention mechanism. These identified interactions serve as edges connecting structural and functional ROIs, forming a unified graph. A graph transformer-based approach leverages this graph for prediction. Experimental results on three datasets demonstrate our model significantly outperforms nine state-of-the-art multi-modal methods in predicting fluid cognition, as well as in classifying Parkinson's disease and schizophrenia. Attention maps show that the visual function engaging with default mode and frontoparietal structures is key to predicting fluid cognition, while diverse functional modules interacting with the subcortical structures are linked to Parkinson's disease. Subcortical and visual-related structures interacting with the functional modules related to sensory and executive function are associated with schizophrenia.

1. Introduction

The brain's functional and structural organizations are characterized by the human connectome as large-scale, complex brain networks [1]. Such connectomes heavily rely on neuroimaging techniques for accurate characterization. Brain functional connectivity (FC) represents temporal relationships among blood-oxygenation-level-dependent signals observed via resting-state functional MRI (rs-fMRI). In contrast, brain structural connectivity (SC) signifies the integrity of white matter pathways, assessed through diffusion MRI (dMRI). Prior research indicates that FC is typically strongest between regions of interest (ROIs) with direct structural connections [2–4]. Changes in the brain structure-function relationship are linked to alterations in cognition [5–7] and the onset of brain disorders [8–10]. Most functional connections, however, arise without a direct structural connection [11–13], suggesting that structure-function interactions extend beyond a straightforward correspondence between paired ROIs across the two modalities [14,15]. Interactions across brain ROIs give rise to intricate mappings between brain function and structure, supporting the brain's large-scale neural communication system [16,17]. Examining these interactions across ROIs can provide deeper insights into the structure, supported by white-matter fiber tracts, that underlie brain function, thereby facilitating the identification of subtle connectivity disruptions. By integrating both structural and functional modalities and moving beyond simple pairwise interactions, analyses of brain disorders and cognition can be significantly enhanced.

The human brain is structured in a modular system [18], with each module comprising several ROIs. ROIs within the same module are

* Corresponding author.
E-mail addresses: jing_xia@zju.edu.cn (J. Xia), yihao.chan@ntu.edu.sg (Y.H. Chan), conghao001@e.ntu.edu.sg (C. Wang), ASJagath@ntu.edu.sg (J.C. Rajapakse).

https://doi.org/10.1016/j.inffus.2026.104426
Received 5 November 2024; Received in revised form 16 April 2026; Accepted 23 April 2026
Available online 24 April 2026

densely connected and strongly coupled internally, while they are more weakly coupled with ROIs in other modules [19]. A popular modular framework organizes the brain into different modules according to functional or structural similarities, such as limbic, frontoparietal, dorsal attention, visual, ventral attention, somatomotor, and default mode modules [20]. Structure-function interactions show homogeneity [21] at the modular level, with ROIs within the same module displaying more similar structure-function coupling neural mechanisms than ROIs across different modules. Thus, using modular-level guidance to capture these complex interactions enhances biological relevance. However, this approach is often overlooked in current studies.

Since brain networks can be represented as graph structures, leveraging graph neural networks (GNNs) for multi-connectome analysis can enhance the exploration of high-level and intricate representations, leading to improved performance [5]. For instance, the Multi-model Graph Convolution Network (M-GCN) [22] applies a filter that acts topologically on the FC matrices, guided by the subject-wise structural connectomes. Multi-view GCN (MV-GCN) [23] employs two separate GCNs for two modalities to learn the modality-specific task-relevant embeddings, which are then combined for prediction. Hypergraph neural network (HGNN) [24] encodes multi-modal connectivity using hyperedges to represent complex relationships. Although these methods combine features from both modalities, they do not consider the interactions between FC and SC.

To address this, Joint-GCN [25] is proposed to fuse structural and functional graphs by adding edges between corresponding regions, followed by GCN to analyze the fused graph. Cross-GNN [26] captures inter-modal dependencies through dynamic graph learning and mutual learning. Moreover, the modality-specific and joint interaction GCN (MSJ-GCN) [27] aims to leverage both modality-specific features and structure-function interactions between corresponding ROIs for prediction tasks. However, these methods primarily focus on one-to-one structure-function coupling between corresponding ROIs and fail to capture the complex interactions across different ROIs. In reality, structure-function coupling in the brain is far more intricate than simple one-to-one interactions. Additionally, modular-level homogeneity provides more biologically meaningful guidance for capturing structure-function interactions. Therefore, a method incorporating modular-level homogeneity and capturing complex interactions across different ROIs is highly desirable.

Building on this, we introduce a modularity-guided interactive attention-based graph neural network (MGIA-GNN) designed to capture complex structure-function interactions, leveraging modular homogeneity to highlight key interactions associated with brain diseases and cognition. In MGIA-GNN, we develop a modularity-guided interactive network to capture interactive strengths across ROIs between two modalities via modularity-specific convolution operations, simulating the coupling homogeneity within modules and heterogeneity across modules. Higher interaction strength indicates a larger coupling strength between function and structure. Next, a modular interactive edge attention model is designed to extract the interactions that are most important for the task based on attention weights. By incorporating the highlighted interactive strengths as edges interlinking all ROIs between the functional and structural graphs, we construct a unified graph that combines the structural and functional connectivity features as well as structure-function interactions across different ROIs. Finally, a graph transformer-based method is applied to this unified graph for the final prediction.

To assess the effectiveness of MGIA-GNN, we perform experiments on three datasets to predict fluid cognition and classify Parkinson's disease (PD) and schizophrenia (SZ), respectively. We compare our framework against two state-of-the-art graph transformer-based models, including HGT [28] and GPS [29], as well as seven state-of-the-art graph-based multi-modal fusion methods, such as M-GCN [22], HGCN [24], Cross-GNN [26], MV-GCN [23], Joint-GCN [25], MSJ-GCN [27], and IMG-GCN [30]. The experiments demonstrate that MGIA-GNN significantly outperforms all nine state-of-the-art graph-based methods on SC and FC. Additionally, MGIA-GNN provides interpretable attention maps that reveal large-scale structure-function interactions, particularly the visual functions interacting with the frontoparietal and default mode structures, as being most salient to predicting fluid cognition. Abnormal interactions between widespread functional modules and subcortical structures are most salient for Parkinson's disease, while subcortical and visual-related structures interacting with the sensory and executive function related functional modules are the most salient for SZ.

Overall, this study has four contributions:

1) We develop an innovative modularity-guided interactive network to capture structure-function interaction strengths across ROIs via modularity-specific convolution operations, which simulate the homogeneity mechanism within modules.
2) We design a modular interactive edge attention model that incorporates cross-attention to identify the most salient interactions for the tasks based on attention weights.
3) We construct a modularity-guided interactive graph that integrates cross-ROI interactions between SC and FC, embedding the extracted coupling strengths as inter-edges linking all ROIs of SC and FC.
4) We reveal, for the first time to the best of our knowledge, the structural-functional interaction patterns associated with functional signal transmission across ROIs for fluid cognition, PD, and SZ.

2. Methodology

The overall architecture, illustrated in Fig. 1, leverages the brain's FC and SC within a modularity-guided interactive network to establish modularity-specific convolution operations, capturing interactions across ROIs. At the same time, a modular interactive edge attention model is applied to identify the most salient interactions. Multiplying the interactions by the edge attention weights, a modularity-guided interactive graph is built, integrating interaction strengths with the SC and FC graphs. This enhanced graph is then fed into a graph transformer-based network for downstream tasks.

2.1. Graph definiations

The brain connectome is naturally represented as a graph, where brain ROIs act as nodes and the connections between these ROIs serve as edges. For each subject, we construct an FC graph $\mathbf{G}^f = (\mathbf{V}^f, \mathbf{A}^f, \mathbf{X}^f)$. $\mathbf{A}^f \in \mathbb{R}^{N\times N}$ is the adjacency matrix, defined as a thresholded FC matrix where the highest 10% of positive edges [31] are retained per node to ensure connection sparsity. The node features are represented by $\mathbf{X}^f = \{\mathbf{x}_1^f, \mathbf{x}_2^f, \ldots, \mathbf{x}_N^f\} \in \mathbb{R}^{N\times D}$, with each $\mathbf{x}_n^f \in \mathbb{R}^{1\times D}$ indicating the functional feature vector of node n with dimension D. Similarly, a SC graph $\mathbf{G}^s = (\mathbf{V}^s, \mathbf{A}^s, \mathbf{X}^s)$ is constructed, where $\mathbf{V}^s$ and $\mathbf{V}^f$ represent the same set of ROIs. The adjacency matrix $\mathbf{A}^s \in \mathbb{R}^{N\times N}$ is derived from dMRI image, and the node feature $\mathbf{X}^s = \{\mathbf{x}_1^s, \mathbf{x}_2^s, \ldots, \mathbf{x}_N^s\} \in \mathbb{R}^{N\times D}$, with each $\mathbf{x}_n^s \in \mathbb{R}^{1\times D}$ indicating the structural feature vector of node n with dimension D. Using the connectivity profile as node features has been reported to be effective in several studies [31,32]. Thus, the feature of each node in $\mathbf{G}^f$ or $\mathbf{G}^s$ is defined by its connectivity with all nodes, making the feature dimension $D = N$.

2.2. Modularity-guided interactive network

In this network, we extract the interactions between function and structure across ROIs, guided by the brain's modular organization. The brain is divided into M modules, with module i containing N_i ROIs, where $\sum_{i=1}^{M} N_i = N$. Given $\mathbf{X}^s$ and $\mathbf{X}^f$, we construct $\mathbf{C}_{ij}$ by concatenating, in a row-wise manner, every functional node feature in module i with every structural node feature in module j. This yields the following modularity-specific feature matrix:

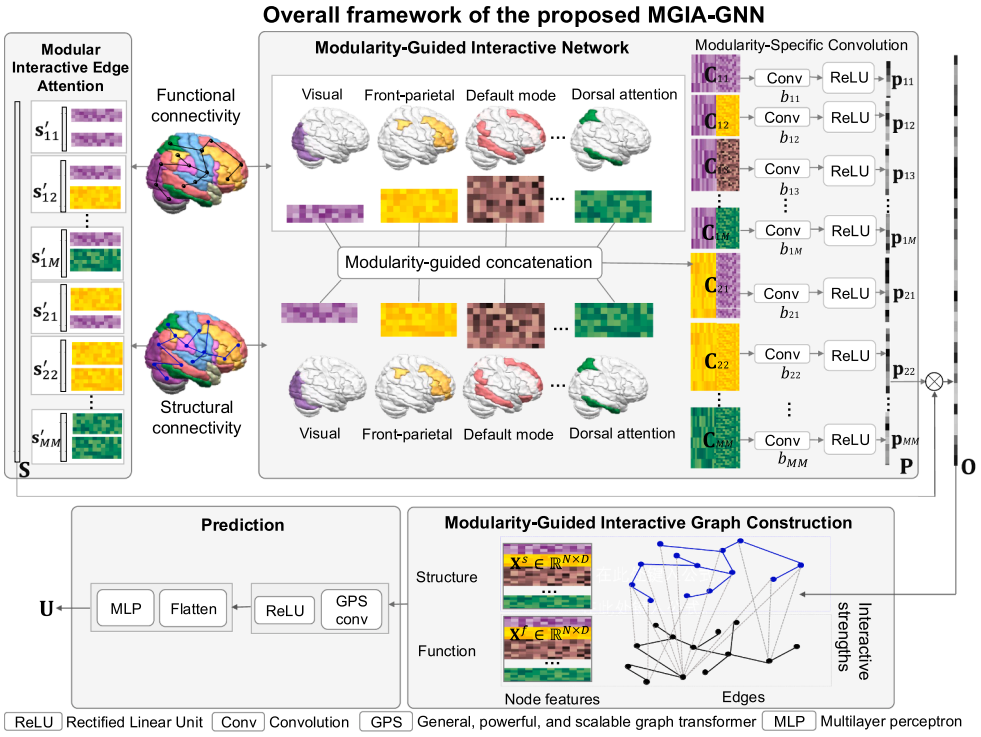

Fig. 1. Overall framework. The MGIA-GNN framework consists of a modularity-guided interactive network leveraging SC and FC to establish modularity-specific convolution operations, capturing interactions across ROIs. Following this, a modular interactive edge attention model identifies the most salient cross-ROI interactions between the two modalities. A modularity-guided interactive graph is then constructed by merging the interaction strengths with the SC and FC graphs. This graph is subsequently processed by a graph transformer-based network (GPS) for downstream prediction tasks.

$$\mathbf{C}_{ij} = \begin{bmatrix} \mathbf{x}^f_{i,1}, & \mathbf{x}^s_{j,1} \\ \mathbf{x}^f_{i,1}, & \mathbf{x}^s_{j,2} \\ \vdots & \vdots \\ \mathbf{x}^f_{i,2}, & \mathbf{x}^s_{j,1} \\ \mathbf{x}^f_{i,2}, & \mathbf{x}^s_{j,2} \\ \vdots & \vdots \\ \mathbf{x}^f_{i,N_i}, & \mathbf{x}^s_{j,N_{j-1}} \\ \mathbf{x}^f_{i,N_i}, & \mathbf{x}^s_{j,N_j} \end{bmatrix} \in \mathbb{R}^{N_i N_j \times 2D}, i, j \in \{1, ., M\}. \tag{1}$$

Here, $\mathbf{x}^f_{i,1} \in \mathbb{R}^{1\times D}$ and $\mathbf{x}^s_{j,1} \in \mathbb{R}^{1\times D}$ refer to the functional node feature and structural node feature of the first ROI in modules i and j, respectively. After concatenating all functional and structural node features, we achieve M^2 modularity-specific features. For each $\mathbf{C}_{ij}$, a modularity-specific convolution operation is defined using a filter b_{ij}, with a filter size of $1 \times 2D$ with stride 1 and no padding, followed by a rectified linear unit (ReLU), as follows:

$$\mathbf{p}_{ij} = ReLU(\mathbf{C}_{ij} * b_{ij}) \in \mathbb{R}^{N_i N_j \times 1}, i, j \in \{1, ., M\}. \tag{2}$$

Here, $*$ represents the convolution operation. Thus, $\mathbf{p}_{ij}$ has dimension $N_i N_j \times 1$, with one scalar corresponding to the interaction between each functional-structural ROI pair. This design is motivated by our hypothesis that ROIs within the same module pair tend to share more similar structure-function coupling patterns than ROIs belonging to different module pairs. Therefore, for each module pair (i, j), the same filter is shared across all ROI pairs within that block, making the operation equivalent to a shared linear projection over each ROI-pair feature within modules i and j. Therefore, the operation is equivalent to a shared linear projection over each ROI-pair feature in module i and module j.

Leveraging this intra-modular coupling homogeneity and inter-modular coupling heterogeneity, we define M^2 convolution filters for M^2 modularity-specific features. The cross-ROI interactions $\mathbf{P}$ for the entire brain are obtained by concatenating the interactions from all modules, as follows,

$$\mathbf{P} = [\mathbf{p}_{11}; \mathbf{p}_{12}; \ldots; \mathbf{p}_{ij}; \ldots; \mathbf{p}_{MM}] \in \mathbb{R}^{N^2 \times 1}, i, j \in \{1, ., M\}. \tag{3}$$

2.3. Modular interactive edge attention model

To explicitly estimate cross-modality ROI-to-ROI interaction between a functional module and a structural module and highlight the most salient interactions of each module in regression or classification, we introduce a modular interactive edge attention model inspired by the cross-attention mechanism [33]. Our attention model consists of M^2 multi-head cross-attention (MMCA) components that generate the attention weights for every pair of functional and structural modules, as shown in Fig. 2. For each MMCA, given the functional node features of all ROIs $\mathbf{L}^f_i = \begin{bmatrix} \mathbf{x}^f_{i,1} \\ \mathbf{x}^f_{i,2} \\ \vdots \\ \mathbf{x}^f_{i,N_i} \end{bmatrix} \in \mathbb{R}^{N_i \times D}$ from the module i, and structural node features of all ROIs $\mathbf{L}^s_j = \begin{bmatrix} \mathbf{x}^s_{j,1} \\ \mathbf{x}^s_{j,2} \\ \vdots \\ \mathbf{x}^s_{j,N_j} \end{bmatrix} \in \mathbb{R}^{N_j \times D}$ from the module j, cross-attention is performed between $\mathbf{L}^f_i$ and $\mathbf{L}^s_j$ as,

$$\mathbf{Q}^h_{ij} = \mathbf{L}^f_i \mathbf{W}^{Q,h}_{ij}, i, j \in \{1, ., M\}, h \in \{1, \ldots, H_1\}, \tag{4}$$

$$\mathbf{K}^h_{ij} = \mathbf{L}^s_j \mathbf{W}^{K,h}_{ij}, i, j \in \{1, ., M\}, h \in \{1, \ldots, H_1\}. \tag{5}$$

Here, $\mathbf{W}^{Q,h}_{ij} \in \mathbb{R}^{D\times(D/H_1)}$ and $\mathbf{W}^{K,h}_{ij} \in \mathbb{R}^{D\times(D/H_1)}$ are learnable parameters for linear projection, and H_1 is the number of attention heads. $\mathbf{Q}^h_{ij} \in \mathbb{R}^{N_i\times(D/H_1)}$ and $\mathbf{K}^h_{ij} \in \mathbb{R}^{N_j\times(D/H_1)}$ represent the query and key in h-th attention head, respectively. To preserve interpretability and keep the parameterization lightweight, the attention weight $\mathbf{s}_{ij}$ is computed

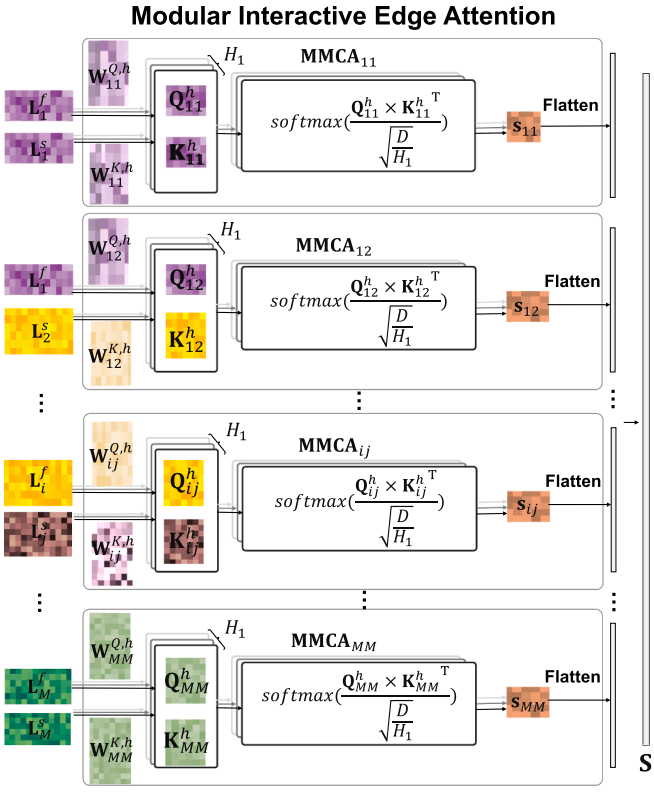

Fig. 2. The architecture of modular interactive edge attention model. For every pair of functional and structural modules, a multi-head cross-attention (MMCA) is performed to capture the modular attention weights.

as:

$$\mathbf{s}_{ij} = average(\mathbf{E}^1_{ij}, \mathbf{E}^2_{ij}, \ldots, \mathbf{E}^{H_1}_{ij}), \tag{6}$$

$$\mathbf{E}^h_{ij} = softmax(\frac{\mathbf{Q}^h_{ij}(\mathbf{K}^h_{ij})^T}{\sqrt{D/H_1}}). \tag{7}$$

$\mathbf{E}^h_{ij}$ denotes the attention weight from h-th attention head, and $\mathbf{s}_{ij} \in \mathbb{R}^{N_i \times N_j}$ indicates the attention weights for the interactions across all structural ROIs in module j and functional ROIs in module i. To ensure the dimensions of the attention weights are consistent with the number of interactions across all ROIs between two modules, we flatten $\mathbf{s}_{ij}$ into $\mathbf{s}'_{ij} \in \mathbb{R}^{N_i N_j \times 1}$. Higher values in the attention weight indicate more important interactions.

The overall attention weight $\mathbf{S}$ is obtained by concatenating the attention weight for each pair of modules, as follows:

$$\mathbf{S} = [\mathbf{s}'_{11}; \mathbf{s}'_{12}; \ldots; \mathbf{s}'_{ij}; \ldots; \mathbf{s}'_{MM}] \in \mathbb{R}^{N^2 \times 1}, i, j \in \{1, ., M\}. \tag{8}$$

The final cross-ROI interaction $\mathbf{O}$ is defined as:

$$\mathbf{O} = \mathbf{S} \otimes \mathbf{P}, \tag{9}$$

where $\otimes$ represents element-wise multiplication. This multiplicative operation makes the resulting interaction matrix more interpretable: a strong final interaction is produced only when both the base interaction strength and the attention weight are high. In this way, the operation enhances interactions that are likely to have a greater impact on the outcome while suppressing those with relatively small contributions. The resulting matrix, $\mathbf{O}$ can then be reshaped into an $N \times N$ matrix, capturing the interactive strengths between the N functional nodes and the N structural nodes.

2.4. Modularity-guided interactive graph construction

Rather than simply concatenating functional and structural graphs, we integrate $\mathbf{G}^f$ and $\mathbf{G}^s$ into one graph $\mathbf{G} = (\mathbf{V}, \mathbf{A}, \mathbf{X})$, by encoding structure-function interactions. The node set of the unified graph $\mathbf{V} = \begin{bmatrix}\mathbf{V}^s \\ \mathbf{V}^f\end{bmatrix}$ has a size of $2N$, resulting in a node feature matrix $\mathbf{X} = \begin{bmatrix}\mathbf{X}^s \\ \mathbf{X}^f\end{bmatrix} \in \mathbb{R}^{2N \times D}$. To link $\mathbf{V}^f$ and $\mathbf{V}^s$ across nodes, we insert the interactive strength matrix $\mathbf{O}$ as $N \times N$ edges that connect all nodes in $\mathbf{V}^f$ and $\mathbf{V}^s$. The resulting adjacency matrix is:

$$\mathbf{A} = \begin{bmatrix}\mathbf{A}^s, \mathbf{O}^T \\ \mathbf{O}, \mathbf{A}^f\end{bmatrix} \in \mathbb{R}^{2N \times 2N}. \tag{10}$$

This integration allows functional and structural profiles across all ROIs to interact, enabling unified graph analysis through a single graph neural network.

2.5. Graph transformer-based method

To fully capture features from the modularity-guided interactive graph, we apply a general, powerful, and scalable graph transformer (GPS) [29] for the final prediction. GPS combines the advantages of transformer methods with GNN-based methods in each layer, effectively learning signal representations over nodes while mitigating the potential overfitting problem in message passing. Specifically, given the unified graph $\mathbf{G}$, a node embedding $\mathbf{X}' \in \mathbb{R}^{2N \times D}$ over nodes is updated by:

$$\mathbf{X}' = \mathrm{GIN}(\mathbf{A}, \mathrm{BN}(\mathbf{X}))), \tag{11}$$

where BN represents the batch normalization operator, and GIN refers to the graph isomorphism network [34]. Compared to traditional GCN, GIN adopts the multi-layer perceptron and introduces the perturbation of central nodes, enhancing its flexibility in capturing complex graph structures. Meanwhile, a transformer-based node embedding $\mathbf{X}'' \in \mathbb{R}^{2N \times D}$ is obtained as:

$$\mathbf{X}'' = concat(\mathbf{E}'^1, \mathbf{E}'^2, \ldots, \mathbf{E}'^{H_2})\mathbf{W}', \tag{12}$$

$$\mathbf{E}'^h = softmax(\frac{\mathbf{Q}'^h(\mathbf{K}'^h)^T}{\sqrt{D/H_2}})\mathbf{F}'^h, h \in \{1, \ldots, H_2\}, \tag{13}$$

where H_2 is the number of attention heads, and $\mathbf{W}' \in \mathbb{R}^{D\times D}$ is a learned linear projection to produce the final node embedding. $\mathbf{Q}'^h$, $\mathbf{K}'^h$, $\mathbf{F}'^h$ are the key, query, and value matrices in h-th attention head, with size $2N \times (D/H_2)$, defined as:

$$\mathbf{Q}'^h = \mathbf{X}\mathbf{W}'^{Q,h}, \tag{14}$$

$$\mathbf{K}'^h = \mathbf{X}\mathbf{W}'^{K,h}, \tag{15}$$

$$\mathbf{F}'^h = \mathbf{X}\mathbf{W}'^{F,h}. \tag{16}$$

Here, $\mathbf{W}'^{Q,h} \in \mathbb{R}^{D\times(D/H_2)}$, $\mathbf{W}'^{K,h} \in \mathbb{R}^{D\times(D/H_2)}$, $\mathbf{W}'^{F,h} \in \mathbb{R}^{D\times(D/H_2)}$ are learnable parameters for the h-th head.

By combining the transformer-based and GIN-based node embeddings, the final node embedding $\mathbf{Z}$ is computed as:

$$\mathbf{Z} = \mathbf{Z}' + MLP_1(\mathbf{Z}'); \tag{17}$$

$$\mathbf{Z}' = \mathrm{BN}(\mathbf{X}' + \mathbf{X}) + \mathrm{BN}(\mathbf{X}'' + \mathbf{X}). \tag{18}$$

MLP_1 is defined as two fully connected layers with dropout and ReLU functions to select the more efficient node embeddings from $\mathbf{Z}'$.

Here, GPS is applied to a unified graph whose node set consists of both structural and functional ROIs. The modality distinction is preserved explicitly by the graph construction itself, and GPS does not treat the two modalities as a completely homogeneous graph input. Rather, it operates on a unified graph with modality-specific structural organization.

2.6. Output layer

Finally, we apply a readout layer that flattens $\mathbf{Z}$, followed by an MLP layer, to generate the predicted outcome. The MLP layer consists of a fully connected layer with dropout and a ReLU function, followed by another fully connected layer to produce the output score, denoted as MLP_2, as follows:

$$\mathbf{U} = MLP_2(Flatten(\mathbf{Z})). \tag{19}$$

Here, because all subjects share the same atlas-defined ROI ordering, flattening preserves node identity explicitly and allows the downstream predictor to use region-specific information directly. In our setting, this is particularly important because biomarker discovery depends not only on global graph summaries but also on retaining the correspondence between each embedding and its anatomical region. For classification, a softmax function is applied after the final fully connected layer to produce the output label. The loss function is defined as cross-entropy for classification tasks, and root mean square error (RMSE) between the true and predicted values for regression tasks.

2.7. Implementation details

All the models were implemented using Python 3.8 and PyTorch 1.12 on an NVIDIA A100 GPU. 5-fold cross-validation was performed to evaluate model performance. We calculated the mean measurement and standard deviation across ten runs. In order to avoid overfitting, we introduced dropout layers after the readout layer and between two fully connected layers in the output layer, with a dropout rate of 0.5. Moreover, we adopted l_2-norm regularization with a weight setting of 1×10^{-3} and employed early stopping to prevent overfitting. The Adam optimizer was used, with the batch size set to 5. For the regression task, the learning rate and number of epochs were set to 1×10^{-3} and 20, respectively. We employed Pearson's correlation coefficient, RMSE, and mean absolute error (MAE) to measure the difference between the real and predicted values of testing samples. For the classification task, the learning rate and number of epochs were set to 1×10^{-4} and 25, respectively. We employed sensitivity, accuracy, and specificity to quantify the classification performance.

3. Experiments and results

3.1. Dataset

We adopted the Human Connectome Project (HCP), Parkinson's Progressive Markers Initiative (PPMI), and Center for Biomedical Research Excellence (COBRE) datasets to assess our model's effectiveness in predicting fluid cognition and classifying Parkinson's disease (PD) and schizophrenia (SZ), respectively. To evaluate our method for fluid cognition prediction, we used data from 839 subjects from the Human Connectome Project (HCP) [35], including corresponding T1w, rs-fMRI, and dMRI images, to evaluate the proposed method. The dMRI and rs-fMRI images were preprocessed using the brain diffusion toolkit within the FSL toolbox [36] and the DPARSF 5.1 advanced toolkit, respectively. Fluid cognition scores, ranging from 87 to 147, were obtained from the phenotype list. For Parkinson's disease classification, we used a dataset of 147 subjects (72 normal controls and 75 patients) from the PPMI dataset [37], including corresponding T1w, dMRI, and rs-fMRI images. For schizophrenia (SZ) classification, we used a dataset of 130 subjects (72 normal controls and 58 patients) from the COBRE dataset [38], including corresponding T1w, rs-fMRI, and dMRI images. Both the PPMI and COBRE datasets were processed using the same pipeline. Specifically, the fMRI images were processed using fMRIPrep [39]. The raw dMRI images were processed via the pipeline in Clinica [40], and SC matrices were obtained by performing probabilistic tractography using the FSL toolbox [36]. Each SC was standardized by log transformation. The FC was constructed by computing the Pearson correlation coefficient between paired ROIs. Both SC and FC networks were constructed with 116 ROIs according to the Anatomical Automatic Labeling (AAL) protocol [41], covering cortical and subcortical regions.

An adapted version of the Yeo 7-network parcellation [20] was applied for defining functional and structural modular systems. This modification introduced an additional subcortical module, including the cerebellum, basal ganglia, thalamus, and vermis, resulting in eight modules ($M = 8$) considered in our study. Specifically, each AAL ROI was represented by its constituent voxels in MNI space, and each of the eight modules was represented by a binary module mask derived from the adapted Yeo-7 partition plus the additional subcortical module. The cortical AAL ROIs were assigned to the Yeo-7 functional modules based on voxel-wise spatial overlap in the same MNI space, whereas the subcortical AAL ROIs were retained as an additional subcortical module. For each cortical AAL ROI and each Yeo module, we computed the overlap by counting the number of voxels simultaneously labeled as ROI in the AAL atlas and as module m in the Yeo atlas. Each cortical ROI was then assigned to the Yeo module with the largest voxel overlap. As a result, our final modular system consisted of seven cortical modules derived from Yeo-7 plus one subcortical module defined from the AAL atlas.

3.2. Hyperparameter setting

In our model, N and D are equal to the number of ROIs, 116. Both the number of attention heads H_1 and H_2 are set as 4 and 4, respectively. The convolution filter has a kernel size of 1×232. The dimension of node embedding of the GPS is set as 116. The number of hidden nodes in the fully connected layers of MLP_1 for GPS is set to 232 and 116, as in [29]. The number of hidden nodes in the fully connected layers of MLP_2 in the output layer is empirically set to 256 and 1, respectively. For the classification task, the number of hidden nodes of the two fully connected layers in the output layer is set to 256 and 2, respectively.

For the baseline methods, the core architectural and training hyperparameters of the baseline models were kept consistent with their original settings. The final output layer is adjusted to suit our application. In the case of HGT and GPS, the number of attention heads was all set as 4, the same as our model. The output layer consists of two fully connected layers with 256 and 1 hidden nodes, respectively, along with a ReLU activation function and a dropout layer. In M-GCN, the layer sizes were

set to 32 channels for the connectome embedding, and 256 for the graph embedding, followed by three fully connected layers with 128, 30, and 1 hidden nodes, respectively, as set in [22]. In HGNN, a two-layer hypergraph network was applied, with the feature dimension of the hidden layer set to 16. In the MV-GCN model, the order of Chebyshev polynomials and output feature dimensions of the graph convolutional layers were set to 30 and 128, respectively, followed by the ReLU activation function, and three fully connected layers with 1024, 116, and 1 hidden nodes, respectively, as described in [23]. In the Joint-GCN, the input and output dimensions of the one-layer GCN were set to 116 and 116, respectively, followed by a scaled exponential linear unit (SELU) activation, mean pooling, and batch-normalization as described in [25]. Three fully connected layers with 3712, 256, and 1 hidden nodes, respectively, were utilized. In Cross-GNN, the size of node embedding of GCN was set as 116, and the number of hidden nodes of three fully connected layers was set as 128, 30, and 1, as detailed in [26]. In the MSJ-GCN, the dimension of node embedding of GCN was set as 64. The kernel size of the convolution filter was set to 1×128. In the IMG-GCN, the dimension of node embedding of GCN was set as 64, and the bottleneck ratio was set as 2. Learning rates were configured according to their original publications. RMSE was used as the loss function for prediction tasks, while cross-entropy was applied for classification. For classification, the final fully connected layer was set to have 2 hidden nodes.

3.3. Comparison with baselines

We comprehensively compared our results against nine state-of-the-art methods that leverage multi-modal features for fluid cognition prediction, as well as PD and SZ classification. These methods include:

1) **Heterogeneity Graph Transformer (HGT)** [28]: HGT is a graph transformer method that applies attention to the features of each node and edge type in a heterogeneous graph.

2) **General, Powerful, Scalable Graph Transformer (GPS)** [29]: GPS combines the strengths of traditional transformers with spatial graph convolution in each layer to effectively learn signal representation over nodes.

3) **Multi-modal GCN (M-GCN)** [22]: M-GCN introduces a filter that acts topologically on the FC matrices, guided by subject-wise structural connectomes.

4) **Hypergraph neural network (HGNN)** [24]: Multiple edges, known as hyperedges, can exist between each pair of connected regions in the hypergraph. HGNN encodes both FC and SC into hyperedges for prediction.

5) **Multi-View GCN (MV-GCN)** [23]: MV-GCN utilizes one GCN for each modality to learn the modality-specific task-relevant embeddings, and then combine them to predict.

6) **Joint-GCN** [25]: Joint-GCN fuses SC and FC networks by adding edges between corresponding regions, allowing the two networks to be analyzed by a single GCN.

7) **Cross-modal GNN (Cross-GNN)** [26]: Cross-GNN captures inter-modal dependencies through dynamic graph learning and mutual learning to fuse multi-modal features and is applied to brain disease classification.

8) **Modality-specific and joint interaction GCN (MSJ-GCN)** [27]: MSJ-GCN incorporates the modality-specific FC and SC features and the interactions between the corresponding ROIs of two modalities for prediction.

9) **Interpretable Modularity-Guided GCN (IMG-GCN)**: Our conference version, IMG-GCN [30], incorporates modular guidance and employs a bottleneck MLP-based attention model to capture structure-function interactions, followed by a GCN for prediction.

For HGT or GPS, we fed FC and SC into a single graph model by stacking them as one graph input, while removing explicit cross-modality edges, so that GPS and HGT operated on a unified graph object rather than on two separately processed graphs. All other methods were implemented following the originally proposed architecture. Tables 1–3 show the comparison results on HCP, PPMI, and COBRE datasets. The mean and standard deviation of ten runs are listed, with the best results in bold. p-value is obtained from the paried *t*-test after FDR correction examining the performance of each method compared to our MGIA-GNN. The results indicate that the HGT and GPS are equal to or outperform multi-modal GCN-based methods that do not utilize the structure-function interactive relationship, such as M-GCN, HGNN, and MV-GCN. This suggests that the graph transformer architecture may capture more effective task-relevant features than simple GCN-based multi-modal methods.

Moreover, multi-modal graph networks that incorporate structure-function interactions, such as Cross-GNN, Joint-GCN, and MSJ-GCN, are equal to or outperform transformer-based methods in most cases, highlighting the importance of structure-function interactions in enhancing performance. Compared with all these approaches that capture the interaction between corresponding ROIs in SC and FC, our proposed method, MGIA-GNN, significantly improves performance across all three datasets, as confirmed by the Wilcoxon signed-rank test ($p < .05$, after FDR correction). For example, MGIA-GNN improves correlation by 9.7% in fluid cognition prediction and enhances accuracy by 2.1% and 4.7% for PD and SZ classification, respectively, compared with MSJ-GCN. We attribute these improvements to the comprehensive capture and utilization of structure-function interactions across different ROIs. Furthermore, our MGIA-GNN significantly outperforms IMG-GCN from our conference version [30], as confirmed by the Wilcoxon signed-rank test in all tasks ($p < .05$). This demonstrates that our modular interactive edge attention model, combined with the GPS model, is more effective for prediction tasks. Note that we additionally report a normalized regression metric, normalized RMSE (NRMSE), to better reflect the prediction error relative to the magnitude of the target score for fluid cognition prediction, indicating that the average prediction error corresponds to approximately 8.02% of the true score at the individual level, as shown in Table S1 in supplementary materials. Moreover, for the classification tasks, we additionally reported ROC-AUC with 95% confidence intervals to better assess model generalizability. For schizophrenia (SZ) classification, the model achieved an AUC of 0.93 (95% CI: 0.88–0.97). For Parkinson's disease classification, the model achieved an AUC of 0.97 (95% CI: 0.95-0.99).

3.4. Ablation study

To evaluate the effectiveness of each component in the proposed MGIA-GNN, we performed ablation experiments on HCP, PPMI, and COBRE datasets. Moreover, the statistical results demonstrate the positive impact of the proposed modules, including the modularity-guided interactive network, the modular interactive edge attention model, and the prediction network, on cognition prediction, as well as the diagnosis of PD and SZ, as shown in Tables 4–6.

1) **w/o MGA**: MGIA-GNN without the modularity-guided interactions and attention module (all values in $\mathbf{O}$ set to 0.1). The value of 0.1 was chosen because it represents a weak connection strength relative to the scale of the original data and serves as a small, nonzero constant that preserves connectivity without imposing overly strong cross-modality coupling.

2) **w/o MG**: MGIA-GNN using a single convolution kernel (not $M \times M$ kernels) applied to all modularity-specific features to capture interactions followed by the attention module.

3) **w/o Att**: MGIA-GNN without the modular interactive edge attention network.

4) **MGI + MA**: Instead of the modular interactive edge attention network, a global cross-attention is used to capture the salient weights based on the entire node embeddings of structure and function.

5) **MGI + MLP**: Instead of the modular interactive edge attention network, a bottleneck-like MLP module with two fully connected layers and one ReLU function is used to learn the salient interactions. It is equal to

Table 1
Comparison of performance with existing methods on HCP for fluid cognition prediction based on FC and SC.

Methods	Correlation	p	RMSE	p	MAE	p
M-GCN	0.26 ± 0.02	9×10^{-3}	12.63 ± 0.68	8×10^{-3}	11.15 ± 0.72	5×10^{-3}
HGNN	0.26 ± 0.03	9×10^{-3}	12.74 ± 1.01	8×10^{-3}	11.27 ± 0.86	5×10^{-3}
MV-GCN	0.27 ± 0.03	9×10^{-3}	12.18 ± 0.86	8×10^{-3}	10.34 ± 0.97	5×10^{-3}
HGT	0.28 ± 0.03	9×10^{-3}	12.15 ± 0.94	8×10^{-3}	10.82 ± 0.63	5×10^{-3}
GPS	0.28 ± 0.01	9×10^{-3}	12.11 ± 0.68	8×10^{-3}	10.65 ± 0.79	5×10^{-3}
Cross-GNN	0.28 ± 0.02	2×10^{-2}	11.95 ± 0.82	8×10^{-3}	10.08 ± 0.53	5×10^{-3}
Joint-GCN	0.29 ± 0.01	2×10^{-2}	11.78 ± 0.89	8×10^{-3}	9.87 ± 0.74	3×10^{-2}
MSJ-GCN	0.31 ± 0.01	2×10^{-2}	11.32 ± 0.77	1×10^{-2}	9.41 ± 0.68	3×10^{-2}
IMG-GCN	0.32 ± 0.02	2×10^{-2}	11.27 ± 0.75	1×10^{-2}	9.36 ± 0.79	3×10^{-2}
Ours	$\mathbf{0.34 \pm 0.01}$	-	$\mathbf{11.01 \pm 0.81}$	-	$\mathbf{9.06 \pm 0.73}$	-

Table 2
Comparison of performance with existing methods on PPMI for PD classification based on FC and SC.

Methods	Accuracy	p	Specificity	p	Sensitivity	p
M-GCN	0.90 ± 0.06	6×10^{-3}	0.91 ± 0.06	4×10^{-3}	0.89 ± 0.05	3×10^{-3}
HGNN	0.82 ± 0.06	6×10^{-3}	0.83 ± 0.08	4×10^{-3}	0.82 ± 0.07	3×10^{-3}
MV-GCN	0.92 ± 0.06	6×10^{-3}	0.92 ± 0.05	4×10^{-3}	0.91 ± 0.08	3×10^{-3}
HGT	0.93 ± 0.04	1×10^{-2}	0.94 ± 0.07	2×10^{-2}	0.92 ± 0.08	2×10^{-2}
GPS	0.92 ± 0.07	1×10^{-2}	0.93 ± 0.06	2×10^{-2}	0.91 ± 0.07	2×10^{-2}
Cross-GNN	0.94 ± 0.07	1×10^{-2}	0.94 ± 0.08	2×10^{-2}	0.93 ± 0.09	2×10^{-2}
Joint-GCN	0.90 ± 0.07	6×10^{-3}	0.91 ± 0.09	4×10^{-3}	0.90 ± 0.07	3×10^{-3}
MSJ-GCN	0.95 ± 0.06	3×10^{-2}	0.96 ± 0.07	4×10^{-2}	0.94 ± 0.05	4×10^{-2}
IMG-GCN	0.95 ± 0.06	3×10^{-2}	0.95 ± 0.06	4×10^{-2}	0.94 ± 0.07	4×10^{-2}
Ours	$\mathbf{0.97 \pm 0.08}$	-	$\mathbf{0.97 \pm 0.09}$	-	$\mathbf{0.96 \pm 0.08}$	-

Table 3
Comparison of performance with existing methods on COBRE for schizophrenia classification based on FC and SC.

Methods	Accuracy	p	Specificity	p	Sensitivity	p
M-GCN	0.82 ± 0.03	7×10^{-3}	0.80 ± 0.03	5×10^{-3}	0.84 ± 0.04	4×10^{-4}
HGNN	0.79 ± 0.02	7×10^{-3}	0.78 ± 0.02	5×10^{-3}	0.80 ± 0.02	1×10^{-3}
MV-GCN	0.82 ± 0.03	2×10^{-2}	0.81 ± 0.06	2×10^{-2}	0.83 ± 0.07	2×10^{-2}
HGT	0.84 ± 0.03	2×10^{-2}	0.82 ± 0.03	2×10^{-2}	0.85 ± 0.04	2×10^{-2}
GPS	0.84 ± 0.08	2×10^{-2}	0.84 ± 0.06	2×10^{-2}	0.85 ± 0.06	2×10^{-2}
Cross-GNN	0.85 ± 0.06	4×10^{-2}	0.84 ± 0.04	3×10^{-2}	0.86 ± 0.05	4×10^{-2}
Joint-GCN	0.84 ± 0.01	2×10^{-2}	0.84 ± 0.01	2×10^{-2}	0.86 ± 0.04	2×10^{-2}
MSJ-GCN	0.86 ± 0.03	2×10^{-2}	0.86 ± 0.04	2×10^{-2}	0.87 ± 0.04	4×10^{-2}
IMG-GCN	0.87 ± 0.03	4×10^{-2}	0.87 ± 0.04	4×10^{-2}	0.88 ± 0.03	4×10^{-2}
Ours	$\mathbf{0.90 \pm 0.06}$	-	$\mathbf{0.89 \pm 0.06}$	-	$\mathbf{0.90 \pm 0.05}$	-

Table 4
Results of ablation study of MGIA-GNN on HCP for fluid cognition prediction.

Methods	Correlation	p	RMSE	p	MAE	p
w/o MGA	0.29 ± 0.02	1×10^{-2}	11.99 ± 0.91	3×10^{-3}	10.04 ± 0.70	4×10^{-3}
w/o MG	0.30 ± 0.02	1×10^{-2}	11.76 ± 0.86	3×10^{-3}	9.71 ± 0.74	4×10^{-3}
w/o Att	0.31 ± 0.05	1×10^{-2}	11.83 ± 0.87	3×10^{-3}	9.56 ± 0.78	4×10^{-3}
MGI + MA	0.32 ± 0.02	2×10^{-2}	11.65 ± 0.80	3×10^{-3}	9.48 ± 0.72	1×10^{-2}
MGI + MLP	0.32 ± 0.03	4×10^{-2}	11.19 ± 0.76	4×10^{-2}	9.23 ± 0.69	1×10^{-2}
Shared conv	0.31 ± 0.02	2×10^{-2}	11.57 ± 0.88	2×10^{-3}	9.35 ± 0.81	1×10^{-2}
MGIA + GCN	0.33 ± 0.02	1×10^{-2}	11.24 ± 0.63	2×10^{-2}	9.20 ± 0.81	2×10^{-2}
MGIA + GAT	0.32 ± 0.02	4×10^{-2}	11.21 ± 0.80	2×10^{-2}	9.28 ± 0.79	2×10^{-2}
MGIA + GIN	0.33 ± 0.02	4×10^{-2}	11.13 ± 0.88	4×10^{-2}	9.29 ± 0.76	2×10^{-2}
MGIA + GPS-18	0.34 ± 0.02	9×10^{-2}	11.18 ± 0.69	4×10^{-2}	9.02 ± 0.58	0.39
MGIA + GPS-8	$\mathbf{0.34 \pm 0.01}$	-	$\mathbf{11.01 \pm 0.81}$	-	$\mathbf{9.06 \pm 0.73}$	-

replacing the original GCN backbone in our previous conference version, IMG-GCN, with GPS while keeping the original interaction setting.

6) **Shared conv**: a single shared convolution filter is used to capture all coupling patterns across module pairs.

7) **MGIA + GCN**: A GCN is used instead of the GPS model for the final prediction, with the node embedding set to 116. It is equal to replacing the original MLP attention module in our previous conference version, IMG-GCN, with a cross-attention mechanism while keeping the original GCN backbone setting.

8) **MGIA + GAT**: A graph attention network (GAT) is used instead of the GPS model for final prediction. GAT uses masked self-attention layers to address shortcomings in prior graph convolution-based methods.

9) **MGIA + GIN**: A graph isomorphism network (GIN) is used instead of the GPS model for final prediction.

10) **MGIA + GPS-18** : This method represents the full proposed MGIA-GNN model using Yeo-17 modules with an additional subcortical module.

Table 5
Results of ablation study of MGIA-GNN on PPMI for PD classification.

Methods	Accuracy	p	Specificity	p	Sensitivity	p
w/o MGA	0.93 ± 0.08	1×10^{-3}	0.94 ± 0.08	3×10^{-3}	0.94 ± 0.06	2×10^{-3}
w/o MG	0.94 ± 0.06	2×10^{-2}	0.95 ± 0.06	3×10^{-3}	0.94 ± 0.07	2×10^{-3}
w/o Att	0.94 ± 0.08	2×10^{-2}	0.94 ± 0.05	2×10^{-2}	0.93 ± 0.07	1×10^{-2}
MGI + MA	0.95 ± 0.06	2×10^{-2}	0.95 ± 0.07	2×10^{-2}	0.94 ± 0.08	1×10^{-2}
MGI + MLP	0.96 ± 0.06	5×10^{-2}	0.96 ± 0.06	2×10^{-2}	0.95 ± 0.10	4×10^{-2}
Shared conv	0.95 ± 0.07	4×10^{-2}	0.94 ± 0.88	2×10^{-2}	0.96 ± 0.81	6×10^{-2}
MGIA + GCN	0.95 ± 0.07	2×10^{-2}	0.96 ± 0.07	6×10^{-2}	0.95 ± 0.09	4×10^{-2}
MGIA + GAT	0.95 ± 0.06	2×10^{-2}	0.96 ± 0.08	4×10^{-2}	0.95 ± 0.07	4×10^{-2}
MGIA + GIN	0.96 ± 0.07	4×10^{-2}	0.96 ± 0.09	4×10^{-2}	0.96 ± 0.08	6×10^{-2}
MGIA + GPS-18	0.96 ± 0.08	0.20	0.97 ± 0.07	0.83	0.95 ± 0.06	0.15
MGIA + GPS-8	**0.97 ± 0.08**	-	**0.97 ± 0.09**	-	**0.96 ± 0.08**	-

Table 6
Results of ablation study of MGIA-GNN on COBRE for SZ classification.

Methods	Accuracy	p	Specificity	p	Sensitivity	p
w/o MGA	0.84 ± 0.02	2×10^{-2}	0.84 ± 0.02	2×10^{-2}	0.85 ± 0.03	2×10^{-2}
w/o MG	0.85 ± 0.02	2×10^{-2}	0.85 ± 0.02	2×10^{-2}	0.86 ± 0.03	2×10^{-2}
w/o Att	0.85 ± 0.05	2×10^{-2}	0.84 ± 0.05	2×10^{-2}	0.86 ± 0.06	2×10^{-2}
MG + MA	0.86 ± 0.02	2×10^{-2}	0.85 ± 0.04	3×10^{-2}	0.87 ± 0.05	2×10^{-2}
MG + MLP	0.87 ± 0.03	3×10^{-2}	0.87 ± 0.03	3×10^{-2}	0.88 ± 0.04	2×10^{-2}
Shared conv	0.87 ± 0.04	3×10^{-2}	0.86 ± 0.05	3×10^{-2}	0.88 ± 0.04	4×10^{-2}
MGIA + GCN	0.88 ± 0.05	4×10^{-2}	0.87 ± 0.07	4×10^{-2}	0.89 ± 0.06	6×10^{-2}
MGIA + GAT	0.87 ± 0.03	4×10^{-2}	0.86 ± 0.06	4×10^{-2}	0.87 ± 0.07	4×10^{-2}
MGIA + GIN	0.88 ± 0.06	4×10^{-2}	0.87 ± 0.04	4×10^{-2}	0.89 ± 0.05	6×10^{-2}
MGIA + GPS-18	0.89 ± 0.04	0.10	0.89 ± 0.04	0.90	0.91 ± 0.03	8×10^{-2}
MGIA + GPS-8	**0.90 ± 0.05**	-	**0.89 ± 0.06**	-	**0.90 ± 0.05**	-

11) **MGIA + GPS-8** (our method): This method represents the full proposed MGIA-GNN model using the current 8 modules. First, a modularity-guided interactive network captures structure-function interactions across ROIs. Then, the modular interactive edge attention module learns the salient interactions for prediction. Finally, GPS is used for the final prediction.

3.4.1. *Effectiveness of the modularity-guided interactive network*

To demonstrate the effectiveness of the modularity-guided interactive network, a comparative analysis using the w/o MGI and w/o MG configurations is conducted. Tables 4–6 show that our proposed modularity-guided interactive network improves prediction performance. Specifically, using single convolutional kernels without modularity guidance (w/o MG) further improves the correlation by 3.4% for fluid cognition prediction and increases the accuracy by 1.1% for PD classification, and 1.2% for SZ classification, compared to setting all interactions to 0.1 (w/o MGA). When employing our modularity-guided interactive (MGI) network, the correlation is enhanced by 13.3% for fluid cognition prediction, while accuracy increases by 3.2% for PD classification and 5.9% for SZ classification, compared to w/o MG. The MGI network incorporates modularity-guided information, simulating the homogeneity of structure-function coupling within modules and heterogeneity across modules, making it more biologically meaningful and leading to better performance compared to those without modularity guidance.

3.4.2. *Effectiveness of the modular interactive edge attention network*

To demonstrate the effectiveness of the modular interactive edge attention network, a comparative analysis with w/o Att, MGI + MA, and MGI + MLP configurations is conducted. Our proposed modular interactive edge attention network enhances the prediction performance, as shown in Tables 4–6. Specifically, compared to the network without any attention module (w/o Att), our modular interactive edge attention network enhances correlation by 9.7% for fluid cognition prediction and increases accuracy by 3.2% for PD classification and 5.9% for SZ classification. When compared to other attention mechanisms, such as global cross-attention (MGI + MA) and a bottleneck-like MLP (MGI + MLP), our modular interactive edge attention improves correlation by 6.3% for fluid cognition prediction and increases accuracy by 2.1% for PD classification and 4.6% and 3.5% for SZ classification, respectively. Moreover, compared to the single convolution (Single conv), our modular convolution improves the correlation while reducing RMSE and MAE for fluid cognition prediction, as well as improving accuracy for Parkinson's disease classification and schizophrenia classification. Our modular interactive edge attention model incorporates a modular-wise cross-attention mechanism, achieving better performance compared to other implementations.

3.4.3. *Effectiveness of the prediction model*

To demonstrate the effectiveness of the prediction model - GPS, a comparative analysis with MGIA + GCN, MGIA + GAT, and MGIA + GIN is conducted. Specifically, compared with all other prediction models, including GCN, GAT, and GIN, the GPS enhances correlation from 3.0% to 6.3% for fluid cognition prediction and enhances the accuracy from 1.0% to 2.1% for PD and 2.3% to 3.5% for SZ classification, respectively. Compared with GCN, GAT, and GIN, the GPS incorporates the transformer module and message-passing network, which can capture more useful information from the node embeddings of the unified graph and enhance the performance in our model.

3.4.4. *Effectiveness of modularity granularity*

To further examine the sensitivity of the framework to the choice of modular granularity, we additionally conducted an ablation study using a finer Yeo-17 network parcellation with an additional subcortex module in place of the original modular partition. The results suggest that the proposed framework remains generally effective under an alternative modular definition, without significant difference between 8 or 18 subnetworks, as highlighted in Tables 4–6, indicating that its performance is not solely tied to one specific parcellation choice.

3.5. *Impact of the hyperparameters*

First, we focused on investigating the effect of the number of attention heads, H_1 and H_2, in the attention and GPS models, respectively, for prediction. Since our number of attention heads should be

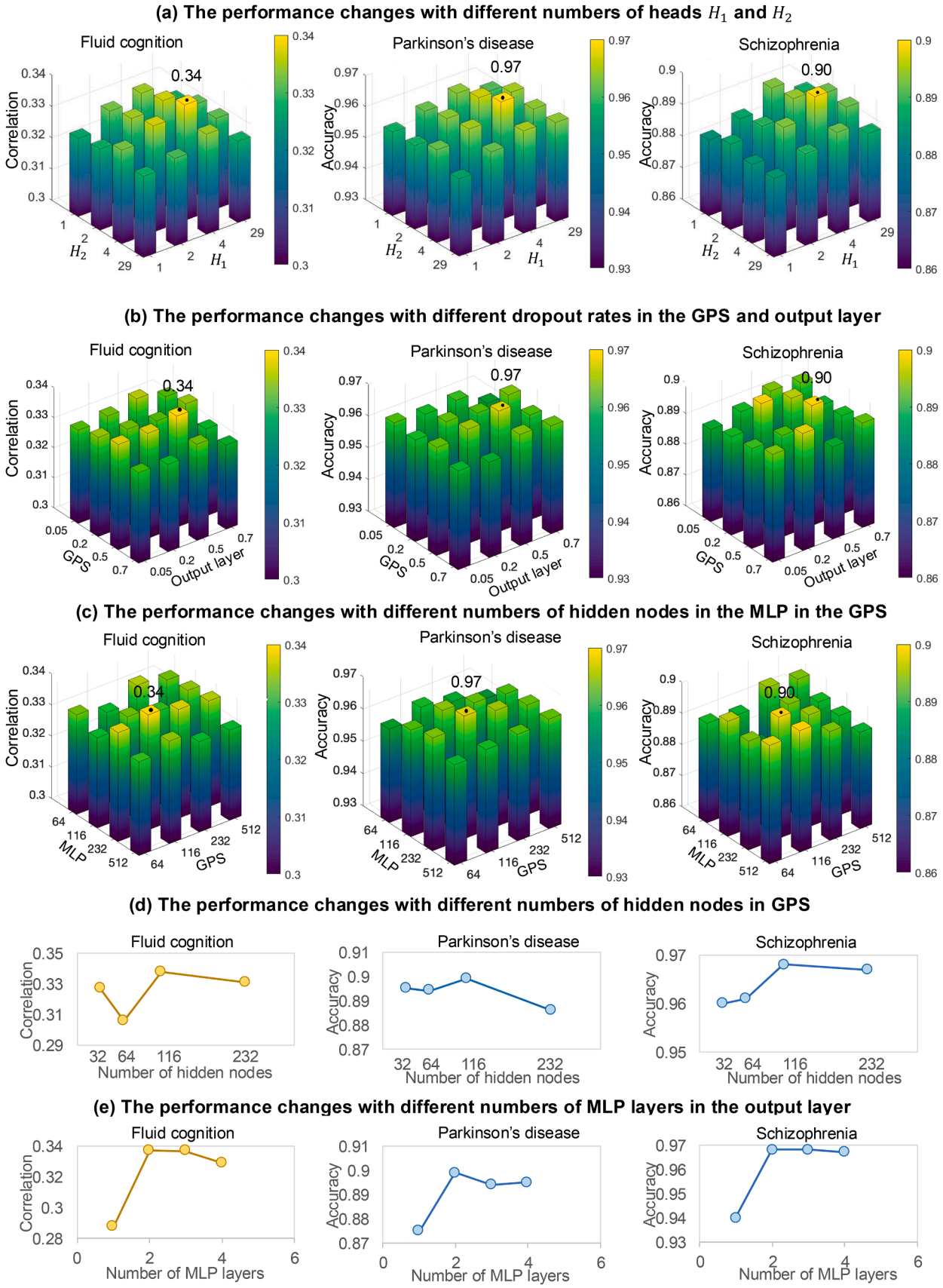

Fig. 3. The performance changes with (a) different numbers of heads results, (b) different dropout rates in the GPS and output layer, (c) different numbers of hidden nodes in the MLP in the GPS, (d) different numbers of hidden nodes in GPS, and (e) different numbers of MLP layers in the output layer, for fluid cognition prediction and both PD and SZ classification. .

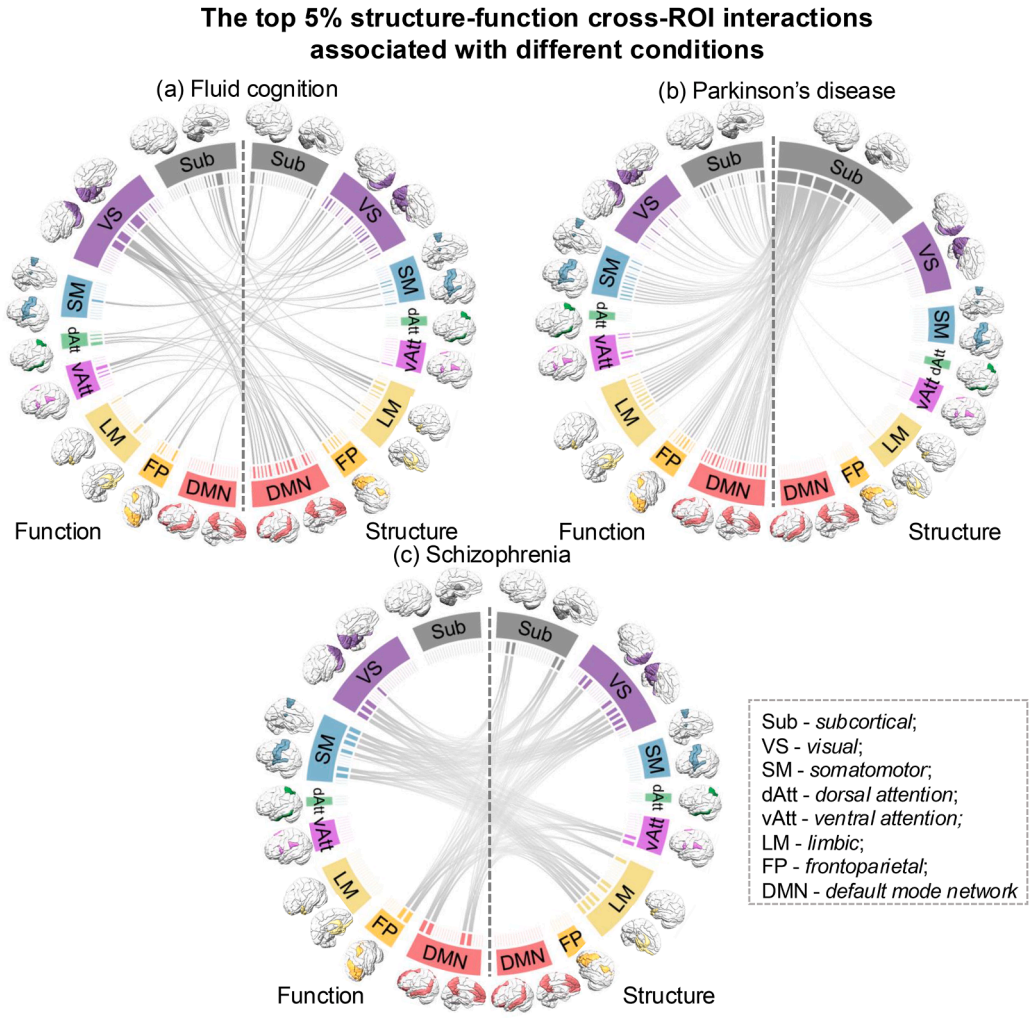

Fig. 4. The modular interactive edge attention model identifies the top 5% of significant structure-function cross-ROI interactions associated with (a) fluid cognition, (b) Parkinson's disease (PD), and (c) schizophrenia (SZ). The attention map reveals that large-scale interactions play a key role in predicting fluid cognition, while alterations of interactions involving subcortical structures and widespread functional modules are most indicative of PD classification. For SZ, interactions between subcortical and visual-related structures with sensory and execution-related functional modules are the most discriminative..

the divisor of the dimension of node embeddings D, we varied H_1 and $H_2 \in \{1, 2, 4, 29\}$, to explore a broad range of head granularities while keeping the feature dimension per head integer-valued. The resulting correlation for fluid cognition prediction and accuracy for PD classification and SZ classification are shown in Fig. 3. It indicates that the optimal hyperparameters are $H_1 = 4$ and $H_2 = 4$, which achieve the best regression performance for fluid cognition and classification performance for PD and SZ.

In addition to the number of attention heads H_1 and H_2, we further report the effects of the dropout rates in the GPS and output layer, the hidden dimensions of the MLP and GPS modules, the hidden dimension in the GPS, and the number of MLP layers in the output layer. As shown in Fig. 3(b), the best-performing dropout combination was a GPS dropout rate of 0.5 and an output-layer dropout rate of 0.5. As shown in Fig. 3(c), the best hidden-dimension combination was 232 hidden units for the MLP in the GPS and 116 hidden units for the GPS representation. Consistently, the one-dimensional analysis in Fig. 3(d) further showed that 116 hidden units were the best choice for the GPS. Finally, Fig. 3(e) showed that using 2 MLP layers in the output layer achieved the best or tied-best performance across tasks, and we therefore adopted 2 layers in the final model.

Moreover, we conducted a sensitivity analysis using multiple sparsity thresholds, including 5%, 10%, and 20% of positive FC edges, as shown in Table S1 in the supplementary materials. The results show that performance remains stable across a reasonable range, 10% gives the best performance, supporting the robustness of the graph construction strategy.

3.6. Model complexity

The modularity-guided interaction network contributes M^2 row-wise filters, while the modular attention block contributes pair-specific linear projections. Our proposed model contains 10.52M trainable parameters, compared with 7.05M for GPS and 6.92M for GCN, as shown in Table 7. In terms of computational cost, using schizophrenia classification as an example, the average training time per epoch is 3.19s for our model, which is higher than GPS and GCN. However, this added complexity is accompanied by a clear performance gain. Our model achieves a mean fold test accuracy of 0.90, compared with 0.85 for GPS and 0.80 for GCN. These results suggest that the proposed model is indeed more computationally demanding than simpler graph baselines, but the improvement in predictive accuracy is substantial, indicating a favourable trade-off between model complexity and performance for this task.

In our study, the graph size remains moderate, and the risk of overfitting is mitigated through dropout, an L2 regularization term in the loss function, and relatively compact hidden layers. We further per-

Table 7
Model complexity, using schizophrenia classification as an example.

Methods	Trainable parameters	Mean accuracy	Epoch time (s)
GCN	6.92M	0.80	0.23
GPS	7.05M	0.85	0.35
Ours	10.52M	0.90	3.19

Table 8
Scalability analysis comparing a 116-node GPS with our 232-node unified graph.

Number of nodes	Mean forward time (ms)	Peak memory (mb)
GPS for 116 nodes	3.49	203.62
GPS for our 232 nodes	6.39	778.59

formed a scalability analysis by comparing GPS on a 116-node graph with our 232-node unified graph. As shown in Table 8, the mean forward time increased from 3.49 ms to 6.39 ms, while peak memory usage increased from 203.62 MB to 778.59 MB. These results indicate that the unified graph introduces additional computational cost, particularly in terms of memory usage. However, the runtime remains within the millisecond range, and the absolute memory usage stays below 1 GB under the current setting. Therefore, although the unified graph is computationally more expensive, it does not impose a prohibitive burden at the graph scale used in this study. More importantly, the main advantage of the unified graph is that it enables explicit cross-modal interaction modeling between FC and SC nodes, which cannot be captured by a single-modality graph. We therefore consider the additional computational cost to be acceptable and justified by its ability to explicitly model cross-modal interactions between FC and SC nodes.

3.7. Discriminative structure-Function interactions

The top 5% discriminative structure-function interactions for fluid cognition prediction and PD and SZ classification, identified by the attention weights $\mathbf{S}$ from the modular interactive edge attention model. To generate the module-level visualization, we mapped each retained ROI pair to its corresponding functional module and structural module according to our modular parcellation. The visualization is constructed by grouping these salient ROI-level interactions based on their module assignments and displaying them within the corresponding functional-module and structural-module sectors, with the module names shown in the outer circle, as shown in Fig. 4. These interactions provide a biologically interpretable view of how structural substrates and functional systems jointly contribute to cognition and disease. Rather than reflecting isolated abnormalities in either SC or FC alone, the highlighted patterns suggest that task-relevant and disease-relevant information is embedded in the coordinated coupling between anatomical pathways and functional communication across modules.

Importantly, the learned convolution filters in the modular interaction module are not interpreted as direct biological quantities at the level of individual parameters. Instead, each module-specific filter can be understood as a data-driven detector of structure-function coupling motifs for a given functional-module and structural-module pair. Because the input features are anatomically indexed connectivity profiles, a strong filter response indicates that a pair of ROIs exhibits a characteristic combination of structural support and functional communication. Therefore, the biological interpretability of the learned filters arises from the spatially localized ROI pairs and module-level circuits that they consistently emphasize, rather than from any single filter coefficient in isolation.

3.7.1. Fluid cognition

Broad structure-function interactions are essential for predicting fluid cognition (see Fig. 4(a)). In particular, the most salient interactions mainly involve the visual functional module together with the limbic, default mode, and frontoparietal structural modules, with additional contributions from attention- and somatomotor-related systems. This pattern is biologically plausible because fluid cognition depends on the coordinated integration of sensory processing, memory-related functions, and higher-order cognitive control.

The frontoparietal and default mode networks are closely related to executive control, flexible reasoning, and internally guided information integration, all of which are central to fluid cognition [42]. Their structural coupling with the visual system may therefore reflect how efficiently externally acquired visual information is transmitted, integrated, and transformed into higher-order cognitive representations. Prior work has also shown that variation in structure-function coupling between higher-order association systems and sensory systems is related to cognitive differences across individuals [43].

The involvement of the limbic structural module further enhances the biological interpretability of our findings. Limbic regions are closely associated with memory formation, salience evaluation, and the integration of sensory input with internal bodily and emotional states [44]. Their interaction with visual functional regions may therefore support the encoding and contextualization of visual experiences, which is particularly relevant to fluid cognition tasks involving visual learning and memory, such as the picture sequence memory test. Consistent with previous studies [7,45], our attention model identifies biologically meaningful structure-function coupling patterns associated with individual differences in fluid cognition.

We further performed subject-level Pearson correlation analyses between the learned module-level structure-function interactive values and cognition scores. The strongest significant association was observed between the functional visual module and the structural frontoparietal module, with $p = 6.24 \times 10^{-5}$. In the provided correlation matrix, the corresponding Pearson correlation coefficient was reported as 0.03. In addition, a significant association was also observed for the functional visual module and structural default mode modules, with $p = .03$. These findings suggest that fluid cognition is particularly related to coupling patterns linking visual processing with higher-order association systems, as well as to within-system coupling in the default mode network.

3.7.2. Parkinson's disease

Abnormal interactions between various functional modules and the subcortical structural module are most salient to classifying PD (see Fig. 4(b)). This finding is biologically meaningful because the subcortical system, particularly the basal ganglia and related thalamic circuits, is central to motor control, action selection, cognitive flexibility, and reward-related processing. PD is classically characterized by progressive structural and microstructural abnormalities in these subcortical regions [46–48].

Compared with recent studies focusing mainly on cortical structure-function coupling [49,50], our findings emphasize that subcortical structural abnormalities may have widespread functional consequences across multiple large-scale brain systems. This is plausible because subcortical nuclei are not isolated relay centers; rather, they participate in distributed cortico-subcortical loops that support motor planning, executive control, and motivational regulation. When the structural integrity of these nuclei and their connecting pathways is compromised, the efficiency and coordination of functional communication with cortical systems may also be altered. Therefore, the highlighted structure-function interactions may reflect how degeneration-related subcortical changes are linked to large-scale disturbances in brain communication in PD. From a biological perspective, our results support the view that PD-related dysfunction is not confined to focal subcortical pathology alone, but is associated with disrupted coupling between subcortical structural substrates and distributed functional systems involved in motor and cognitive processing.

Moreover, we performed module-level two-sample t-tests on the summed structure-function interactive values between controls and patients. Significant interactive alterations were mainly concentrated

around the subcortical module. Specifically, for the functional subcortical module, significantly reduced interactive values in PD was observed with the structural subcortical module, with the group mean difference computed as control minus patient equal to 0.07 ($p = 6.29 \times 10^{-7}$), the structural visual module, with the group mean difference computed as control minus patient equal to 0.06 ($p = 2.4 \times 10^{-22}$), the structural limbic module, with the group mean difference computed as control minus patient equal to 0.05 ($p = 8.17 \times 10^{-18}$), the structural frontoparietal module, with the group mean difference computed as control minus patient equal to 0.8 ($p = 1.92 \times 10^{-21}$). For the structural subcortical module, significantly reduced interactive values in PD was also observed with the functional visual module, with the group mean difference computed as control minus patient equal to 1.1 ($p = 2.67 \times 10^{-28}$), the functional ventral attention module, with the group mean difference computed as control minus patient equal to 0.02 ($p = .024$), and the functional limbic module, with the group mean difference computed as control minus patient equal to 0.04 ($p = .04$), These results indicate that PD-related abnormalities are centered on subcortical interactions with visual, limbic, and higher-order association systems.

3.7.3. Schizophrenia

Abnormal structure-function interactions involving subcortical, visual, and limbic structural modules interacting with somatomotor, visual, frontoparietal, and default mode functional modules—crucial for sensory and executive functions—are highly distinguishing for schizophrenia classification (see Fig. 4(c)). This pattern is biologically interpretable because schizophrenia is widely understood as a disorder of distributed brain dysconnectivity, affecting both sensory processing and higher-order cognitive integration.

The involvement of visual and somatomotor functional modules suggests that schizophrenia-related abnormalities are not limited to executive systems, but also extend to early sensory and perceptual processing pathways. This is consistent with evidence that schizophrenia is associated with altered connectivity in occipital and sensory-related regions, as well as impairments in perceptual organization and sensory integration [51,52]. At the same time, the frontoparietal and default mode functional modules are closely related to cognitive control, self-referential processing, and large-scale network integration, which are frequently disrupted in schizophrenia [53]. The corresponding structural modules further support this interpretation. Alterations in subcortical and limbic structural systems may influence the anatomical support for efficient communication between sensory and higher-order functional systems. In this sense, the identified interactions suggest that schizophrenia may involve coordinated abnormalities in the structural substrates underlying sensory-limbic processing and the functional networks responsible for executive and integrative cognition. Rather than indicating simple changes in SC or FC alone, these biomarkers point to altered structure-function coupling across circuits that bridge perceptual, salience-related, and executive processes.

Moreover, we performed the same module-level two-sample t-tests between controls and SZ patients. Significant interaction alterations were concentrated in subcortical, visual, limbic, and frontoparietal systems. Specifically, for the functional subcortical module, significant group differences were observed with the structural visual module, with the group mean difference computed as control minus patient equal to 0.2 ($p = 7.01 \times 10^{-6}$), the structural frontoparietal module, with the group mean difference computed as control minus patient equal to 0.7 ($p = .038$), and the structural default mode module, with the group mean difference computed as control minus patient equal to 0.08 ($p = .005$). Additional significant differences were observed between the functional limbic and structural visual modules, with the group mean difference computed as control minus patient equal to 0.3 ($p = .04$), and between the functional frontoparietal and structural frontoparietal modules, with the group mean difference computed as control minus patient equal to 0.07($p = .01$). These results suggest that SZ-related abnormalities are more distributed and involve both reduced and increased interactive values, consistent with dysregulated structure-function coordination across perceptual, salience-related, and executive systems.

4. Discussion

Our proposed method offers three key advantages. First, it demonstrates promising performance on three public datasets, outperforming nine state-of-the-art multi-modal methods. This robustness is essential for clinical applications. Second, our method identifies salient structure-function connections across different ROIs, unlike existing methods that only capture interactions between corresponding ROIs. The cross-ROI structure-function coupling provided by our approach can be further utilized to uncover causal relationships between structural and functional alterations in brain disorders. Third, our method introduces a novel approach by integrating biological characteristics into the model. It can be easily adapted to incorporate other brain features, such as node hubs.

Our motivation for adopting a modular template, such as a modified Yeo-7 template, was to provide a common and biologically meaningful reference for organizing cross-modality interactions at the group level. More importantly, our assumption is not that all ROIs within the same module must exhibit identical interaction values, nor that all subjects share the same exact modular architecture. Rather, we assume that ROI-level interactions within the same functional-module and structural-module block are more likely to follow a similar coupling mechanism, whereas interactions from different module pairs may exhibit stronger heterogeneity. This is the reason why, in our model, ROI pairs within the same module pair share the same convolutional parameters, while different module pairs are assigned different convolution filters. In this sense, the modularity assumption serves primarily as a structured inductive bias for interaction learning, rather than as a claim of fixed biological uniformity across all subjects.

In our framework, each node feature is defined as the ROI-level connectivity profile to all other ROIs, resulting in a feature dimension of $D = N = 116$. We adopted this representation because it preserves full connectome-level context for each ROI and provides a comprehensive description of its connectivity pattern within the brain network. Although this choice is higher-dimensional than many manually designed node descriptors, a dimensionality of 116 remains moderate in the context of connectome analysis. Retaining the full connectivity profile helps preserve interpretability and avoids discarding potentially informative connections at the input stage.

Moreover, the proposed interaction module is implemented using a $1 \times 2D$ convolution over concatenated structural-functional features. Its role in our framework is not merely that of a lightweight linear projection. Rather, it serves as the first step in a structured interaction-learning pipeline that is sufficiently expressive to capture complex cross-modality interactions in the current setting. Specifically, the use of module-specific filters allows different module pairs to learn heterogeneous interaction patterns, while the subsequent ReLU nonlinearity introduces nonlinear transformation of the projected features. Recently, transformer-based methods have demonstrated efficient performance [54–58]. Inspired by these, our learned interactions are then further refined by the modular cross-attention mechanism, which adaptively reweights interaction signals in a task-dependent manner, and by the downstream GPS model, which propagates and transforms the interaction-aware graph representation at a higher level. Therefore, the effective expressive power of the interaction modeling arises from the composition of module-specific projection, nonlinear activation, adaptive attention reweighting, and graph-based feature transformation, rather than from the convolution operation alone. To further examine whether a more sophisticated interaction module is necessary, we additionally implemented a two-layer MLP-based variant to replace the $1 \times 2D$ convolution for each module pair, as shown in Table S3 in the supplementary materials. Using schizophrenia classification as an example, this MLP-based design did not achieve better performance than the original convolution-based module, despite substantially increasing

the number of interaction parameters from 14,848 to 3474496. These results suggest that the proposed convolution-based interaction module is not only parameter-efficient but also already sufficiently effective for capturing task-relevant complex interactions in our framework, providing a favorable balance between representational power and model complexity.

Note that differences in preprocessing pipelines between HCP and the other datasets may introduce a potential confounding factor. However, the aim of this study was not to perform direct head-to-head comparisons across datasets, but to use them as independent benchmarks to assess whether the proposed framework can operate robustly under distinct neuroimaging settings. Notably, despite the heterogeneous preprocessing regimes, the proposed method consistently achieved the best performance within each dataset, supporting its robustness and generalizability across different data conditions. At the same time, the identified biomarkers should be interpreted primarily within each dataset, as preprocessing differences may affect their direct comparability across datasets.

There are some limitations that need to be addressed. First, while SC and FC may share certain large-scale organizational patterns, they are governed by partly distinct principles: SC primarily reflects relatively stable anatomical wiring, whereas FC captures dynamic and context-dependent functional coordination. Therefore, using the same Yeo-7 modular system for both modalities in the present study is a simplification rather than an ideal representation of their true organization. We adopted a shared modular template mainly to provide a common reference space for modeling cross-modality interactions and to avoid introducing an additional source of instability, particularly given the limited availability of a well-established SC-specific modular system at the same scale. However, we fully acknowledge that this design may bias the learned interactions toward an artificially aligned structure-function correspondence, potentially underestimating modality-specific modular properties and limiting the biological specificity of the resulting interpretations. In future work, we aim to establish distinct modular systems for SC and FC, respectively, based on their respective connectivity patterns, and utilize these systems to better capture structure-function coupling. Second, the current framework still relies on a population-level fixed modular template, it may introduce bias when subject-specific or disorder-altered modular structure deviates from the predefined partition, and jointly learned modular structures, in order to better capture individual variability and disease-related reorganization.

5. Conclusion

We proposed a modularity-guided interactive attention-based structure-function coupling network to predict cognition and brain diseases. Our method significantly outperforms nine state-of-the-art methods, including two graph transformer-based methods and seven multimodal fusion methods. The attention model's salient maps reveal distinctive, biologically meaningful structure-function interactions across ROIs related to SZ, PD, and fluid cognition. Our interpretable results are likely to improve the understanding of PD and SZ pathophysiology and promote the identification of new structure-function coupling markers to diagnose the diseases.

CRediT authorship contribution statement

Jing Xia: Writing – original draft, Visualization, Validation, Methodology, Conceptualization; **Yi Hao Chan:** Writing – review & editing, Software, Formal analysis, Data curation; **Deepank Girish:** Writing – review & editing; **Conghao Wang:** Writing – review & editing; **Jagath C. Rajapakse:** Funding acquisition, Conceptualization, Project administration, Supervision.

Data availability

All the data used in the manuscript are in public domain.

Declaration of competing interest

We declare that there are no conflicts of interest to report.

Acknowledgment

Research supported by AcRF Tier-2 grant MOE T2EP20121-0003 and Tier-1 grant RG15/24 of Ministry of Education, Singapore.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.inffus.2026.104426

References

[1] O. Sporns, G. Tononi, R. Kötter, The human connectome: a structural description of the human brain, PLoS Comput. Biol. 1 (4) (2005) e42.
[2] Y. Li, R. Dong, X. Liu, Y. Yuan, Y. Kong, Neurobridge: bridging functional and structural brain networks via neural coupling and consistency-guided dynamic graph learning, Med. Image Anal. 110 (2026) 103993.
[3] G.L. Baum, Z. Cui, D.R. Roalf, R. Ciric, R.F. Betzel, B. Larsen, M. Cieslak, P.A. Cook, C.H. Xia, T.M. Moore, et al., Development of structure–function coupling in human brain networks during youth, Proc. Natl. Acad. Sci. 117 (1) (2020) 771–778.
[4] E. Dhamala, K.W. Jamison, A. Jaywant, S. Dennis, A. Kuceyeski, Distinct functional and structural connections predict crystallised and fluid cognition in healthy adults, Hum. Brain Mapp. 42 (10) (2021) 3102–3118.
[5] S. Li, Q. Zhu, C. Tian, W. Shao, D. Zhang, Interpretable dynamic brain network analysis with functional and structural priors, IEEE Trans. Med. Imaging 44 (2025) 4878–4889.
[6] X. Dong, Q. Li, X. Wang, Y. He, D. Zeng, L. Chu, K. Zhao, S. Li, How brain structure–function decoupling supports individual cognition and its molecular mechanism, Hum. Brain Mapp. 45 (2) (2024) e26575.
[7] Z. Gu, K.W. Jamison, M.R. Sabuncu, A. Kuceyeski, Heritability and interindividual variability of regional structure-function coupling, Nat. Commun. 12 (1) (2021) 4894.
[8] A. Droby, S. Nosatzki, Y. Edry, A. Thaler, N. Giladi, A. Mirelman, I. Maidan, The interplay between structural and functional connectivity in early stage Parkinson's disease patients, J. Neurol. Sci. 442 (2022) 120452.
[9] L. Wang, Y. Xu, H. Zhang, L. Chen, W. Dou, X. Yin, J. Ye, Y.-C. Chen, Optimization of structural connectomes and scaled patterns of structural-functional decoupling in Parkinson's disease, NeuroImage 284 (2023) 120450.
[10] R. Davoodi, M. Mehrabbeik, S. Jafari, M. Perc, Elevated EEG fractal dimension in Parkinson's during timing tasks, Chaos Interdiscip. J. Nonlinear Sci. 35 (6) (2025).
[11] G. Qu, Z. Zhou, V.D. Calhoun, A. Zhang, Y.-P. Wang, Integrated brain connectivity analysis with fMRI, DTI, and sMRI powered by interpretable graph neural networks, Med. Image Anal. 103 (2025) 103570.
[12] L.E. Suárez, R.D. Markello, R.F. Betzel, B. Mišić, Linking structure and function in macroscale brain networks, Trends Cogn. Sci. 24 (4) (2020) 302–315.
[13] M. Gosak, M. Milojević, M. Duh, K. Skok, M. Perc, Networks behind the morphology and structural design of living systems, Phys. Life Rev. 41 (2022) 1–21.
[14] Z. Wei, T. Dan, J. Ding, P.J. Laurienti, G. Wu, NeuroDetour: a neural pathway transformer for uncovering structural-functional coupling mechanisms in human connectome, Med. Image Anal. 109 (2026) 103931.
[15] T. Sarwar, Y. Tian, B.T.T. Yeo, K. Ramamohanarao, A. Zalesky, Structure-function coupling in the human connectome: a machine learning approach, NeuroImage 226 (2021) 117609.
[16] S. Majhi, S. Ghosh, P.K. Pal, S. Pal, T.K. Pal, D. Ghosh, J. Završnik, M. Perc, Patterns of neuronal synchrony in higher-order networks, Phys. Life Rev. 52 (2025) 144–170.
[17] C. Seguin, M.P. van den Heuvel, A. Zalesky, Navigation of brain networks, Proc. Natl. Acad. Sci. 115 (24) (2018) 6297–6302.
[18] O. Sporns, R.F. Betzel, Modular brain networks, Annu. Rev. Psychol. 67 (1) (2016) 613–640.
[19] M.A. Bertolero, B.T.T. Yeo, M. D'Esposito, The modular and integrative functional architecture of the human brain, Proc. Natl. Acad. Sci. 112 (49) (2015) E6798–E6807.
[20] B.T.T. Yeo, F.M. Krienen, J. Sepulcre, M.R. Sabuncu, D. Lashkari, M. Hollinshead, J.L. Roffman, J.W. Smoller, L. Zöllei, J.R. Polimeni, B. Fischl, H. Liu, R.L. Buckner, The organization of the human cerebral cortex estimated by intrinsic functional connectivity, J. Neurophysiol. 106 (3) (2011) 1125–1165.
[21] H.-J. Park, K. Friston, Structural and functional brain networks: from connections to cognition, Science 342 (6158) (2013) 1238411.
[22] N.S. Dsouza, M.B. Nebel, D. Crocetti, J. Robinson, S. Mostofsky, A. Venkataraman, M-GCN: A multi-modal graph convolutional network to integrate functional and structural connectomics data to predict multidimensional phenotypic characterizations, in: Medical Imaging with Deep Learning, 2021, pp. 119–130.

[23] X. Zhang, L. He, K. Chen, Y. Luo, J. Zhou, F. Wang, Multi-view graph convolutional network and its applications on neuroimage analysis for Parkinson's disease, in: AMIA Annual Symposium Proceedings, vol. 2018, American Medical Informatics Association, 2018, p. 1147.
[24] Y. Feng, H. You, Z. Zhang, R. Ji, Y. Gao, Hypergraph neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 3558–3565.
[25] Y. Li, Q. Wei, E. Adeli, K.M. Pohl, Q. Zhao, Joint graph convolution for analyzing brain structural and functional connectome, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2022, pp. 231–240.
[26] Y. Yang, C. Ye, X. Guo, T. Wu, Y. Xiang, T. Ma, Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning, IEEE Trans. Med. Imaging 43 (1) (2024) 108–212.
[27] J. Xia, Y.H. Chan, D. Girish, J.C. Rajapakse, Brain structure-function interaction network for fluid cognition prediction, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023, pp. 1706–1710.
[28] Z. Hu, Y. Dong, K. Wang, Y. Sun, Heterogeneous graph transformer, in: Proceedings of the Web Conference 2020, 2020, pp. 2704–2710.
[29] L. Rampášek, M. Galkin, V.P. Dwivedi, A.T. Luu, G. Wolf, D. Beaini, Recipe for a general, powerful, scalable graph transformer, Adv. Neural Inf. Process. Syst. 35 (2022) 14501–14515.
[30] J. Xia, Y.H. Chan, D. Girish, J.C. Rajapakse, IMG-GCN: Interpretable modularity-guided structure-function interactions learning for brain cognition and disorder analysis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2024.
[31] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L.H. Staib, P. Ventola, J.S. Duncan, BrainGNN: interpretable brain graph neural network for fMRI analysis, Med. Image Anal. 74 (2021) 102233.
[32] H. Cui, W. Dai, Y. Zhu, X. Kan, A.A.C. Gu, J. Lukemire, L. Zhan, L. He, Y. Guo, C. Yang, BrainGB: a benchmark for brain network analysis with graph neural networks, IEEE Trans. Med. Imaging 42 (2) (2023) 493–506.
[33] C.-F.R. Chen, Q. Fan, R. Panda, CrossViT: cross-attention multi-scale vision transformer for image classification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 357–366.
[34] K. Xu, W. Hu, J. Leskovec, S. Jegelka, How powerful are graph neural networks?, 2018, arXiv:1810.00826.
[35] D.C. Van Essen, S.M. Smith, D.M. Barch, T.E.J. Behrens, E. Yacoub, K. Ugurbil, W.-M. HCP Consortium, The WU-Minn human connectome project: an overview, NeuroImage 80 (2013) 62–79.
[36] M. Jenkinson, C.F. Beckmann, T.E.J. Behrens, M.W. Woolrich, S.M. Smith, FSL, NeuroImage 62 (2) (2012) 782–790.
[37] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, et al., The parkinson progression marker initiative (PPMI), Prog. Neurobiol. 95 (4) (2011) 629–635.
[38] C.J. Aine, H.J. Bockholt, J.R. Bustillo, J.M. Cañive, A. Caprihan, C. Gasparovic, F.M. Hanlon, et al., Multimodal neuroimaging in schizophrenia: description and dissemination, Neuroinformatics 15 (2017) 343–364.
[39] O. Esteban, C.J. Markiewicz, R.W. Blair, C.A. Moodie, A.I. Isik, A. Erramuzpe, J.D. Kent, et al., fMRIPrep: a robust preprocessing pipeline for functional MRI, Nat. Methods 16 (1) (2019) 111–116.
[40] A. Routier, N. Burgos, M. Díaz, M. Bacci, S. Bottani, O. El-Rifai, S. Fontanella, et al., Clinica: an open-source software platform for reproducible clinical neuroscience studies, Front. Neuroinforma. 15 (2021) 689675.
[41] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, NeuroImage 15 (1) (2002) 273–289.
[42] M.L. Rosen, D. Amso, K.A. McLaughlin, The role of the visual association cortex in scaffolding prefrontal cortex development: a novel mechanism linking socioeconomic status and executive function, Dev. Cogn. Neurosci. 39 (2019) 100699.
[43] G. Feng, Y. Wang, W. Huang, H. Chen, J. Cheng, N. Shu, Spatial and temporal pattern of structure-function coupling of human brain connectome with development, Elife 13 (2024) RP93325.
[44] V.S. Ramachandran, Encyclopedia of the Human Brain, Elsevier, 2002.
[45] G.L. Baum, Z. Cui, D.R. Roalf, R. Ciric, R.F. Betzel, B. Larsen, M. Cieslak, et al., Development of structure–function coupling in human brain networks during youth, Proc. Natl. Acad. Sci. 117 (1) (2020) 771–778.
[46] L. Wang, C. Zhou, W. Cheng, E.T. Rolls, P. Huang, N. Ma, Y. Liu, et al., Dopamine depletion and subcortical dysfunction disrupt cortical synchronization and metastability affecting cognitive function in Parkinson's disease, Hum. Brain Mapp. 43 (5) (2022) 1598–1610.
[47] R.A. Menke, J. Scholz, K.L. Miller, S. Deoni, S. Jbabdi, P.M. Matthews, M. Zarei, MRI characteristics of the substantia nigra in Parkinson's disease: a combined quantitative T1 and DTI study, NeuroImage 47 (2) (2009) 435–441.
[48] X. Bai, C. Zhou, T. Guo, X. Guan, J. Wu, X. Liu, T. Gao, et al., Progressive microstructural alterations in subcortical nuclei in Parkinson's disease: a diffusion magnetic resonance imaging study, Park. Relat. Disord. 88 (2021) 82–89.
[49] A. Zarkali, P. McColgan, L.-A. Leyland, A.J. Lees, G. Rees, R.S. Weil, Organisational and neuromodulatory underpinnings of structural-functional connectivity decoupling in patients with Parkinson's disease, Commun. Biol. 4 (1) (2021) 86.
[50] M. Wang, C. Tan, Q. Shen, S. Cai, Q. Liu, H. Liao, Altered functional-structural coupling may predict Parkinson's patient's depression, Brain Struct. Funct. 229 (4) (2024) 897–907.
[51] E.A. Reavis, J. Lee, L.L. Altshuler, M.S. Cohen, S.A. Engel, D.C. Glahn, A.M. Jimenez, et al., Structural and functional connectivity of visual cortex in schizophrenia and bipolar disorder: a graph-theoretic analysis, Schizophr. Bull. Open 1 (1) (2020) sgaa056.
[52] H. Tohid, M. Faizan, U. Faizan, Alterations of the occipital lobe in schizophrenia, Neurosci. J. 20 (3) (2015) 213–224.
[53] L. Cocchi, I.H. Harding, A. Lord, C. Pantelis, M. Yucel, A. Zalesky, Disruption of structure–function coupling in the schizophrenia connectome, NeuroImage Clin. 4 (2014) 779–787.
[54] Z. Hou, J. Liu, Z. Shao, Q. Ma, W. Liu, Machine learning innovations in renewable energy systems with integrated NRBO-TXAD for enhanced wind speed forecasting accuracy, Electronics 14 (12) (2025) 2329.
[55] Z. Hou, J. Liu, S. Yu, Enhanced analog circuit fault diagnosis via continuous wavelet transform and dual-stream convolutional fusion, Sci. Rep. 15 (1) (2025) 19828.
[56] Z. Hou, B. Wang, J. Liu, Y. He, Y. Yao, Physics-inspired time-frequency feature extraction and lightweight neural network for power quality disturbance classification, Front. Phys. 13 (2025) 1616367.
[57] J. Liu, Z. Duan, X. Hu, J. Zhong, Y. Yin, Detracking autoencoding conditional generative adversarial network: improved generative adversarial network method for tabular missing value imputation, Entropy 26 (5) (2024) 402.
[58] J. Liu, Z. Hou, T. Yin, Short-term power load forecast using OOA optimized bidirectional long short-term memory network with spectral attention for the frequency domain, Energy Rep. 12 (2024) 4891–4908.