

BrainRMA: ROI-Preserving Multiscale Attention for Brain Network Analysis

Anonymous submission

Abstract

Functional brain interactions are studied non-invasively from blood-oxygen-level-dependent (BOLD) signals using resting-state functional magnetic resonance imaging (rs-fMRI). This Functional brain-network analysis requires modeling both long-range dependencies among regions of interest (ROIs) and cluster-level organization among functionally related regions. Existing graph neural networks are limited by local message passing, while graph Transformers capture long-range interactions but often ignore the community structures of ROIs. Conversely, generic clustering or pooling methods introduce cluster structure but they are not specialized for brain networks where collapsing multiple ROIs into one representation, loses region-specific information. To bridge this gap in brain network analysis, we propose BrainRMA, an ROI-preserving multiscale attention framework. BrainRMA models functional brain graphs through two complementary views: a long-range view that captures distributed ROI dependencies and a cluster-aware view that represents ROI-to-cluster and cluster-level interactions using training-derived functional clusters. The two views are then fused at the ROI level, followed by DEC-based pooling then graph classification. We evaluate our proposed method on two publicly available datasets: ABIDE for Autism Spectrum Disorder (ASD) classification and OSF for brain age classification. Our model achieves remarkable accuracies of 82.35% (ABIDE) and 93.75% (OSF), surpassing current state-of-the-art (SOTA) baselines by 3.35% (ABIDE) and 6.25% (OSF), respectively. We further demonstrate that BrainRMA captures meaningful ROI-to-cluster patterns across temporal, frontal, parietal, default-mode, sensorimotor, subcortical, and cerebellar systems, supporting ROI-preserving multiscale attention as an effective and interpretable strategy for functional brain graph classification. We will publicly release the source code and trained models.

1 Introduction

Resting-state functional magnetic resonance imaging (rs-fMRI) provides a non-invasive means of characterizing functional interactions across the brain. By measuring blood-oxygen-level-dependent (BOLD) signals from anatomically defined regions of interest (ROIs), rs-fMRI enables the construction of functional connectivity (FC) networks in which nodes represent regions of brain and edges quantify statistical dependencies between their activity patterns. These networks provide a structured representation for studying neurological and developmental differences. However, learning discrimi-

This is an anonymized submission for review purposes only. Distribution, citation, or public sharing of this manuscript is strictly prohibited. Copyright and publication details will appear in the final version if accepted.

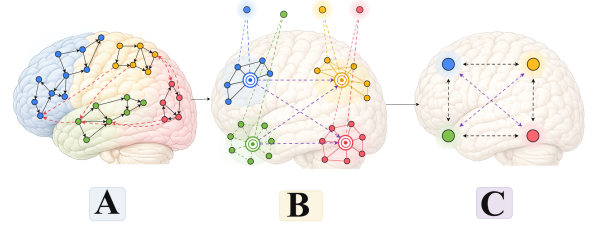


Figure 1: **Multiscale communication in BrainRMA.** (a) Long-range interactions capture dependencies between distant ROIs. (b) ROI-to-cluster attention introduces functional-module context while preserving ROI-level representations. (c) Cluster-level interactions summarize communication among functional modules, connecting global ROI reasoning with modular brain organization.

native representations from FC remains challenging because brain connectivity is high dimensional, noisy, and distributed across interacting regional and system-level organizations.

Graph neural networks (GNNs) (Kipf and Welling 2017; Veličković et al. 2018) have substantially advanced FC analysis by explicitly exploiting the relational structure of brain networks. Brain-specific architectures can identify informative regions, learn task-dependent connectivity patterns, and aggregate evidence over functional neighborhoods (Li et al. 2021; Kan et al. 2022b). Nevertheless, conventional message passing remains inherently local. Information exchanged between distant functional systems must traverse several intermediate ROIs, making long-range dependencies difficult to preserve and increasing the risk of over-smoothing as the network becomes deeper.

Graph Transformers provide an attractive alternative because self-attention allows direct interaction between arbitrary ROIs. BrainNetTF represents each ROI through its whole-brain connectivity profile and learns global attention-based representations (Kan et al. 2022b). More recently, ALTER introduced correlation-biased random walks to construct adaptive multi-hop structural encodings before global self-attention (Yu et al. 2024). These developments demonstrate that brain-graph classification benefits from capturing dependencies beyond immediate functional neighbors.

However, unrestricted global interaction alone doesn't

fully reflect the organization of functional brain networks. A large body of network-neuroscience research shows that the brain is organized into coherent functional systems rather than as an undifferentiated collection of pairwise connections (Rubinov and Sporns 2010; Power et al. 2011). Densely interacting regions form specialized modules that support functional segregation, while selective inter-module connections and connector regions enable coordinated whole-brain integration (Gallos, Makse, and Sigman 2012; Bertolero, Yeo, and D’Esposito 2015). Thus, modularity and long-range communication are complementary mechanisms.

This distinction is particularly relevant for clinical and developmental analysis. Autism spectrum disorder (ASD) has been associated with altered segregation and integration across large-scale functional systems, including changes involving default-mode, sensory, and higher-order association networks (Rudie et al. 2013; Yang et al. 2023). Brain-network organization also changes throughout development, progressing from relatively local connectivity patterns toward more distributed and differentiated functional systems (Fair et al. 2009; Richardson et al. 2018). These findings suggest that discriminative information may appear simultaneously in within-module organization and in interaction that span multiple functional systems.

Existing learning architecture capture these scales partially. Global Transformers provide global communication, but typically process the brain as a flat sequence of tokens. Consequently, it must also infer both fine-grained local dependencies and higher-order functional organization from limited supervision. Community-aware models introduce mesoscopic structure more explicitly, but commonly rely on either predefined partitions or ROI-to-community compression. Fixed communities may not match the functional organization of a particular group, whereas merging multiple ROIs removes local distinction precisely when system level context is introduced (Bannadabhavi et al. 2023; Yang et al. 2024).

Node-to-Cluster Attention (N2C-Attn) provides a promising mechanism for addressing this information loss by allowing individual ROIs to interact with cluster-level representations (Huang et al. 2024). However, applying a general node-to-cluster operator raises three important questions: (i) How can a common set of functional modules be estimated consistently across subjects, (ii) How should modules of unequal reliability be weighted?, and (iii) How should modular communication be coordinated with long-range interactions that cross module boundaries? This leads to the central question:

How can a brain-graph exploit group-level functional modules while preserving individual ROI representations and retaining whole brain communication?

To address this question, we propose BrainRMA, a multiscale framework that jointly models global ROI interactions and functional module-aware communication through two complementary paths. For each evaluation split, we construct a group-level FC graph using only the training subjects, retain its strongest connections, and partition it into functional modules using anatomical and data-driven clustering

methods, like METIS (Karypis and Kumar 1998), spectral clustering (Ng, Jordan, and Weiss 2002), Louvain clustering (Blondel et al. 2008), Yeo-network grouping (Yeo et al. 2011), DMO-N (Tsitsulin et al. 2023), etc. This provides a shared group-level organization.

The module-aware pathway first refines interactions within each module and then forms module representations. Each ROI attends to these representations through kernelized node-to-cluster attention, gaining system-level context. To account for variations in partition quality, each module is weighted by its internal-to-external connectivity ratio, while a learnable residual gate regulates the overall module-aware update. In parallel, the global pathway captures long range ROI-to-ROI dependencies.

Key Contributions are as follows:

- We propose BrainRMA, a novel dual-pathway attention framework that concurrently models global, long-range ROI dependencies and the local functional modular communication without compressing or over-smoothing individual regional features.
- We introduce an interpretable ROI-to-cluster attention mechanism, providing neuroscientifically aligned region-cluster relationships.
- We evaluate BrainRMA on the ABIDE dataset for Autism Spectrum Disorder (ASD) classification and the OSF dataset for brain age prediction. Extensive experiments show that our method achieves state-of-the-art accuracies of 82.35% and 93.75%, surpassing current baselines by absolute margins of 3.35% and 6.25%, respectively.
- We are making the source code and data splits publicly available for the community to support further research.

2 Related Work

Resting-state functional connectivity exhibits a multiscale organization in which specialized modules interact to support whole-brain function. Accordingly, brain-graph methods can be grouped by how they model local connectivity, global ROI interactions, and community-level structure.

Message-Passing Models for Brain Graphs Graph neural networks aggregate information from connectivity-defined neighborhoods. GCN, GraphSAGE, and GAT use normalized, learned, or attention-based neighborhood aggregation, respectively (Kipf and Welling 2017; Hamilton 2017; Veličković et al. 2018). Brain-specific models further introduce connectome filters, ROI-selection pooling, graph-design strategies, and task-adaptive connectivity learning (Kawahara et al. 2017; Li et al. 2021; Cui et al. 2022; Kan et al. 2022a). Although effective for local patterns, distant functional interactions require repeated propagation in these methods, while modular structure is generally used only during graph-level pooling.

Global Interaction in Brain Transformers Graph Transformers directly model distant node interactions using positional, spectral, shortest-path, or hybrid local-global structural biases (Dwivedi and Bresson 2020; Kreuzer et al. 2021; Ying et al. 2021; Wu et al. 2021). BrainNetTF, GBT, and ALTER adapt this principle to connectomes through

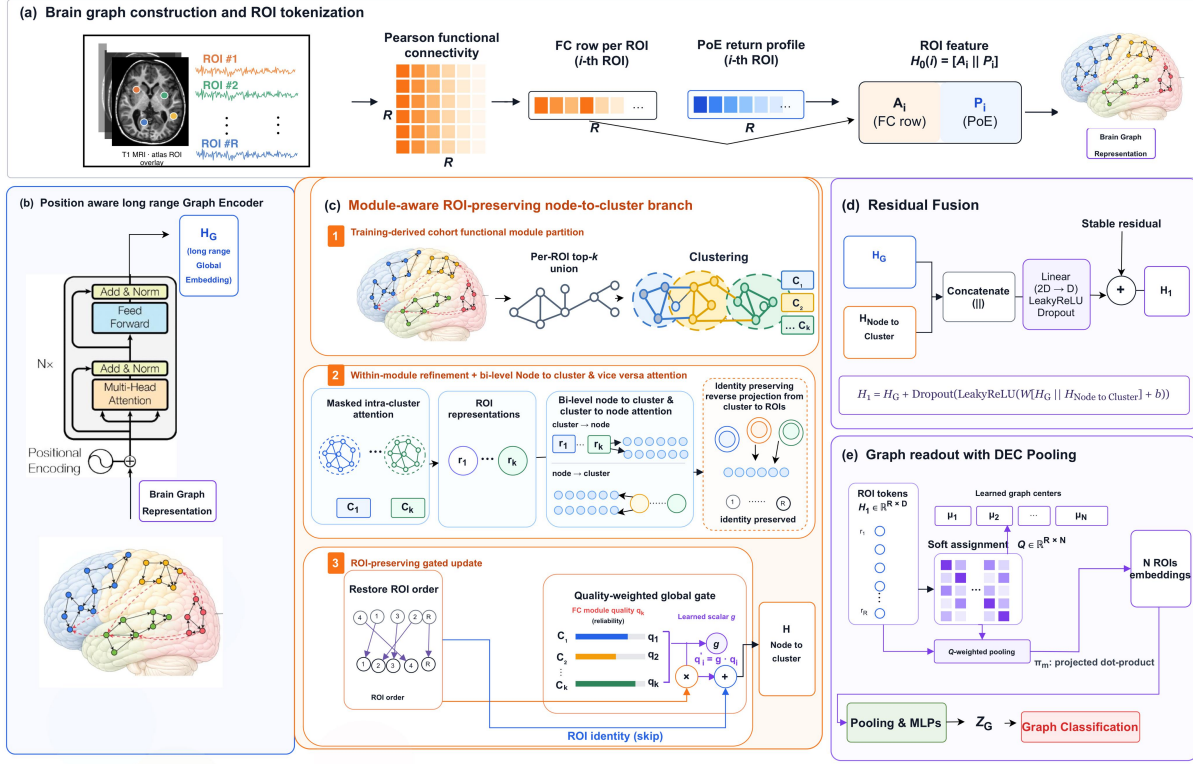


Figure 2: **Overview of the proposed framework.** (a) A subject-specific brain graph is constructed from Pearson functional connectivity, with each ROI represented by its connectivity row and positional encoding (PoE) profile. (b) A position-aware long-range graph encoder produces ROI-aligned global representations H_G . (c) The cluster aware branch partitions the training-derived graph into functional modules, refines ROI features through intra-cluster and bidirectional ROI-cluster attention, and restores the original ROI order using a quality-weighted residual update. (d) Residual Fusion combines the global and cluster-aware ROI representations to obtain H_1 . (e) DEC-inspired pooling assigns ROI tokens to learned centers and produces a graph-level representation for subject-level classification.

whole-brain ROI profiles, representation regularization, and correlation-biased multi-hop signatures (Kan et al. 2022b; Peng et al. 2024; Yu et al. 2024). These methods capture distributed dependencies but generally operate on flat ROI sequences, requiring functional modules to be learned implicitly from subject-level supervision.

Community-Aware and Hierarchical Models Community-aware methods introduce intermediate functional structure using fixed graph partitions or learned assignments. TC-BrainTF, for example, learns community representations through prompt-based centroids and deep embedded clustering (Yang et al. 2024; Xie, Girshick, and Farhadi 2016). However, fixed communities may poorly reflect group-specific organization, whereas learned community tokens can compress distinct ROI information. Node-to-Cluster Attention preserves node representations while incorporating cluster-level context through multiplicative or learnable linear combinations of node- and cluster-level kernels (Huang et al. 2024). Nevertheless, its general formulation does not define how functional modules

should be derived from multi-subject connectivity, weighted by structural reliability, or combined with unrestricted global reasoning.

Our framework addresses these aspects through training-derived functional modules, reliability-aware node-to-cluster attention, and a parallel long range dependency pathway. The two ROI-aligned representations are integrated through learnable residual fusion, enabling modular and global interactions to complement each other without collapsing regional information.

3 Methodology

Preliminaries For each subject s , let $\mathbf{B}^{(s)} \in \mathbb{R}^{N \times T_s}$ denote the preprocessed resting-state fMRI time series, where N is the number of ROIs and T_s is the number of time points. Functional connectivity between ROIs i and j is computed using Pearson correlation, $C_{ij}^{(s)} = \text{corr}(\mathbf{b}_i^{(s)}, \mathbf{b}_j^{(s)})$. The resulting matrix $\mathbf{C}^{(s)} \in \mathbb{R}^{N \times N}$ represents the subject-specific functional network. In our implementation, the Pear-

son functional-connectivity matrix is used both as the node feature matrix $\mathbf{X}^{(s)}$ and as the graph adjacency $\mathbf{A}^{(s)}$:

$$\mathbf{X}^{(s)} = \mathbf{C}^{(s)}, \quad \mathbf{A}^{(s)} = \mathbf{C}^{(s)}. \quad (1)$$

Thus, each ROI is represented by its whole-brain connectivity profile, and ROI-to-ROI edges are given by the same subject-specific functional-connectivity matrix. The initial ROI token is formed by concatenating this connectivity feature with the random-walk relative positional encoding:

$$\mathbf{Z}^{(s,0)} = [\mathbf{X}^{(s)} \parallel \mathbf{P}^{(s)}] \in \mathbb{R}^{N \times (N+d)}. \quad (2)$$

Each ROI is represented by its whole-brain connectivity profile. Therefore, the node feature matrix is defined as $\mathbf{X}^{(s)} = \mathbf{C}^{(s)}$. To introduce structural information, the connectivity features are combined with random-walk relative positional encodings. Let $\mathbf{P}^{(s)} \in \mathbb{R}^{N \times d}$ denote the positional encoding. The initial ROI representation is

$$\mathbf{Z}^{(s,0)} = [\mathbf{X}^{(s)} \parallel \mathbf{P}^{(s)}] \in \mathbb{R}^{N \times d_h}, \quad d_h = N + d, \quad (3)$$

where $\parallel$ denotes feature-wise concatenation. The model uses a long-range graph encoder to capture distributed ROI interactions and a node-to-cluster mechanism to introduce functional-module information.

Problem Formulation A subject-specific functional brain graph is represented as $\mathcal{G}^{(s)} = (\mathcal{V}, \mathcal{E}^{(s)}, \mathbf{X}^{(s)}, \mathbf{A}^{(s)})$, where $\mathcal{V} = \{v_i\}_{i=1}^N$ is the set of ROIs, $\mathcal{E}^{(s)}$ is the subject-specific edge set, $\mathbf{X}^{(s)}$ is the ROI feature matrix, and $\mathbf{A}^{(s)}$ is the functional-connectivity adjacency matrix. Given a dataset $\mathcal{D} = \{(\mathcal{G}^{(s)}, y^{(s)})\}_{s=1}^L$, the objective is to learn a graph representation $\mathbf{z}_G^{(s)}$ and a classifier f_θ such that

$$\hat{\mathbf{y}}^{(s)} = f_\theta(\mathbf{z}_G^{(s)}). \quad (4)$$

The proposed framework learns this representation by jointly modeling global ROI interactions and functional-module organization.

Long range and Cluster Representation Learning For each evaluation split, a common ROI partition is constructed using only the training subjects $\mathcal{T}$. The group-level functional-connectivity matrix is computed as

$$G_{ij} = \frac{1}{|\mathcal{T}|} \sum_{s \in \mathcal{T}} |C_{ij}^{(s)}|, \quad G_{ii} = 0. \quad (5)$$

The correlations allow strong positive and negative functional associations to contribute to the group graph. Each ROI retains its k_{top} strongest connections, producing a sparse group topology. The ROIs are then clustered into approximately balanced functional modules while reducing cross-module connections. Given a target average patch size p , where p denotes the desired average number of ROIs inside one functional module/cluster, the number of functional modules is set to $K = \lceil N/p \rceil$, where N is the number of ROIs. The resulting partition $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_K\}$ is fixed for validation and testing to avoid information leakage.

The initial ROI features are processed through two complementary pathways. The long range pathway captures distributed interactions among all ROIs:

$$\mathbf{H}^G = f_{\text{global}}(\mathbf{Z}^0, \mathbf{A}) \in \mathbb{R}^{N \times d_h}. \quad (6)$$

The clustering pathway captures structured interactions within and across functional groups. For each cluster $\mathcal{C}_k$, the corresponding ROI features are processed by a shared intra-ROI Transformer, written as $\tilde{\mathbf{Z}}_k = \text{EnC}_{\text{intra}}(\mathbf{Z}_k^0)$. The representation of ROI k is obtained by averaging its refined ROI features:

$$\mathbf{r}_k = \frac{1}{|\mathcal{C}_k|} \sum_{i \in \mathcal{C}_k} \tilde{\mathbf{z}}_i. \quad (7)$$

The ROI representations are stacked as $\mathbf{R}_c = [\mathbf{r}_1, \dots, \mathbf{r}_K]^\top$. We use two node-to-cluster attention kernels. The first is the linear kernel variant, denoted BrainRMA-L, which uses a positive feature map to compute ROI-to-cluster attention efficiently. The second is the dot-product kernel variant, denoted BrainRMA-S, which uses softmax attention between each ROI token and the cluster representations. Both variants preserve ROI-level tokens and use clusters only as contextual modules. For the linear attention variant, the positive kernel is $\phi(\mathbf{x}) = \text{ELU}(\mathbf{x}) + 1$, ϵ acting as normalization factor and the node-to-cluster context is

$$\mathbf{C}^L = \frac{\phi(\tilde{\mathbf{Z}}) [\phi(\mathbf{R}_c)^\top \mathbf{R}_c]}{\phi(\tilde{\mathbf{Z}}) [\phi(\mathbf{R}_c)^\top \mathbf{1}_K] + \epsilon}. \quad (8)$$

For the dot-product variant, the attention coefficient is defined as $a_{ik} = \text{softmax}_k(\tilde{\mathbf{z}}_i^\top \mathbf{r}_k / \sqrt{d_h})$, and the context of ROI i is $\mathbf{c}_i = \sum_{k=1}^K a_{ik} \mathbf{r}_k$.

For either variant, the cluster context is integrated with the refined ROI representation as

$$\mathbf{u}_i = \tilde{\mathbf{z}}_i + \text{ReLU}([\tilde{\mathbf{z}}_i \parallel \mathbf{c}_i] \mathbf{W}_c + \mathbf{b}_c). \quad (9)$$

After ROI-to-cluster interaction, the updated ROI tokens are restored to the original atlas ROI order, as shown in Fig. 2. This is done using the cluster membership map. If ROI i belongs to cluster $\mathcal{C}_k$, its updated representation $\mathbf{u}_i$ is written back to the i -th position of the output tensor $\mathbf{H}_i^N = \mathbf{u}_i$ where $i = 1, \dots, N$. This index-based reverse projection does not pool, average, or merge ROIs. Instead, it restores each cluster-processed token to its original brain-region index. From a neurobiological perspective, this is important because each ROI corresponds to a specific anatomical or functional brain region. Therefore, the model can use functional modules as contextual brain systems while still preserving region-specific information for each ROI. Unlike conventional graph-pooling or cluster-coarsening methods that collapse multiple ROIs into a single module token, our reverse projection keeps every ROI anatomically traceable. This keeps the node-to-cluster branch output ROI-aligned with the global branch, enabling ROI-wise fusion and interpretation at both individual brain-region and functional-module levels.

The reliability of each cluster is estimated using the ratio between its mean internal and external connectivity. Let I_k

Table 1: Performance comparison of the proposed method with recent learnable-network and graph-transformer baselines on ABIDE and OSF using the Craddock-200 (Craddock et al. 2013) and AAL (Tzourio-Mazoyer et al. 2002) atlases, respectively. The superscript [#] denotes the corresponding BrainCKT-enhanced variant (Meng et al. 2026), which is the current SOTA. **Bold** ones are the best and the underlined ones are the second best metrics.

Model Type	Method	Venue	ABIDE				OSF			
			ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE
Learnable Network	GCN (Kipf and Welling 2017)	ICLR 2017	58.39	61.45	41.43	76.12	75.00	76.00	28.57	88.00
	GraphSAGE (Hamilton 2017)	NeurIPS 2017	60.58	65.03	65.71	55.22	78.12	80.57	71.43	80.00
	GAT (Veličković et al. 2018)	ICLR 2018	59.12	60.83	65.71	52.24	62.50	81.14	85.71	56.00
	BrainGNN (Li et al. 2021)	MedIA 2021	59.40	62.40	36.70	70.70	45.22	51.79	55.10	44.35
	BrainGB (Cui et al. 2022)	IEEE TMI 2023	63.64	69.71	64.05	61.31	46.70	49.38	52.71	49.32
	BrainNetCNN (Kawahara et al. 2017)	NeuroImage 2017	67.80	74.98	64.32	71.04	46.70	49.38	52.71	49.32
	FBNETGNN (Kan et al. 2022a)	MIDL 2022	68.01	75.67	65.71	62.94	46.70	49.38	52.71	49.32
Graph Transformer	SAN (Kreuzer et al. 2021)	NeurIPS 2021	65.30	71.30	55.40	68.30	<u>87.50</u>	89.71	71.43	<u>92.00</u>
	VanillaTF (Vaswani et al. 2017)	NeurIPS 2017	65.20	76.40	66.40	71.10	68.75	84.57	<u>85.70</u>	64.00
	Graphormer (Ying et al. 2021)	NeurIPS 2021	60.84	63.57	78.72	36.72	55.00	58.88	43.33	52.51
	TC-BrainTF (Yang et al. 2024)	arXiv 2024	69.70	77.71	69.12	70.15	61.11	62.08	45.34	67.85
	BrainNETTF (Kan et al. 2022b)	NeurIPS 2022	71.00	80.02	72.50	69.30	60.00	62.59	66.67	55.56
	Com-BrainTF (Bannadabhavi et al. 2023)	MICCAI 2023	79.60	72.50	80.10	65.70	63.33	75.00	71.80	68.13
	ALTER (Yu et al. 2024)	NeurIPS 2024	77.00	82.80	77.40	76.60	66.67	74.55	81.80	55.15
	BrainCKT+BrainNETTF [#]	AAAI 2026	75.00	81.25	76.60	73.59	62.21	68.52	44.44	83.33
	BrainCKT+Com-BrainTF [#]	AAAI 2026	78.00	83.11	<u>81.81</u>	73.33	62.79	76.55	75.32	67.78
	BrainCKT+ALTER [#]	AAAI 2026	79.00	<u>83.87</u>	78.51	<u>78.31</u>	73.33	81.48	62.50	60.00
Ours	BrainRMA-S (Softmax Kernel Attn.)		<u>81.02</u>	83.24	80.00	82.09	93.75	<u>90.29</u>	85.71	96.00
	BrainRMA-L (Linear Kernel Attn.)		82.35	85.36	82.61	82.09	93.75	90.86	85.71	96.00

and E_k denote these quantities. The quality score is defined as $s_k = I_k / (E_k + \epsilon_q)$ and is normalized across modules to obtain q_k . Here, q_k denotes the normalized reliability score of cluster $\mathcal{C}_k$. Larger q_k indicates that the ROIs inside the cluster have stronger within-cluster connectivity relative to their external connectivity, and therefore the cluster $\mathcal{C}_k$ is more functionally coherent.

Let $\kappa(i)$ denote the cluster containing ROI i , and let $g = \sigma(\ell_g)$ be a learnable global gate. The cluster-aware representation is

$$\mathbf{h}_i^M = \tilde{\mathbf{z}}_i + gq_{\kappa(i)}(\mathbf{u}_i - \tilde{\mathbf{z}}_i). \quad (10)$$

The quality coefficient reduces the influence of weak clusters, while the learnable gate controls the overall contribution of module information.

Fusion and Graph Prediction The global and cluster representations preserve the same ROI ordering and have shape $\mathbf{H}^G, \mathbf{H}^M \in \mathbb{R}^{N \times d_h}$. They are concatenated as $\mathbf{H}^{\text{cat}} = [\mathbf{H}^G \| \mathbf{H}^M]$. A nonlinear correction is computed as $\Delta \mathbf{H} = \text{Dropout}(\text{LeakyReLU}(\mathbf{H}^{\text{cat}} \mathbf{W}_f + \mathbf{b}_f))$, and the fused representation is

$$\mathbf{H}^F = \mathbf{H}^G + \Delta \mathbf{H}. \quad (11)$$

The global representation acts as the residual reference, while the learned correction introduces complementary cluster-aware information.

The fused ROI representations are summarized using center-assignment pooling. For ROI i and pooling center μ_m , the assignment score is $s_{im}^{\text{pool}} = (\mathbf{h}_i^{F\top} \mu_m) / \|\mu_m\|_2$, and the normalized assignment is $\pi_{im} = \exp(s_{im}^{\text{pool}}) / \sum_{r=1}^M \exp(s_{ir}^{\text{pool}})$. The pooled component associated with center m is

$$\mathbf{p}_m = \frac{\sum_{i=1}^N \pi_{im} \mathbf{h}_i^F}{\sum_{i=1}^N \pi_{im}}. \quad (12)$$

The pooled components are stacked into $\mathbf{H}^{\text{pool}}$, projected as $\mathbf{R} = \text{LeakyReLU}(\mathbf{H}^{\text{pool}} \mathbf{W}_r + \mathbf{b}_r)$, and flattened to obtain $\mathbf{z}_G = \text{vec}(\mathbf{R})$. The final prediction is $\hat{\mathbf{y}} = \text{MLP}(\mathbf{z}_G)$.

The model is optimized using cross-entropy loss:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{|\mathcal{B}|} \sum_{s \in \mathcal{B}} \sum_{c=1}^C y_{sc} \log \left(\text{softmax} \left(\hat{\mathbf{y}}^{(s)} \right)_c \right). \quad (13)$$

4 Experiments

Experimental Settings

Datasets We evaluate the proposed model on two well-known public brain-network analysis datasets of different scales: the Autism Brain Imaging Data Exchange (ABIDE) (Di Martino et al. 2014) and the Theory-of-Mind task dataset available through the Open Science Framework (OSF) (Richardson et al. 2018). Brain regions are defined using the Craddock atlas for ABIDE (Craddock et al. 2013) and the AAL atlas for OSF (Tzourio-Mazoyer et al. 2002). The dataset descriptions are provided in supplementary.

Implementation Details All experiments were implemented in PyTorch on a single NVIDIA RTX A4000 GPU. We used the Craddock-200 atlas for ABIDE and the 116-region AAL atlas for OSF. Subjects were stratified into training, validation, and test sets in a 70:10:20 ratio following the work (Meng et al. 2026). The group-level functional connectivity graph was constructed using only the training

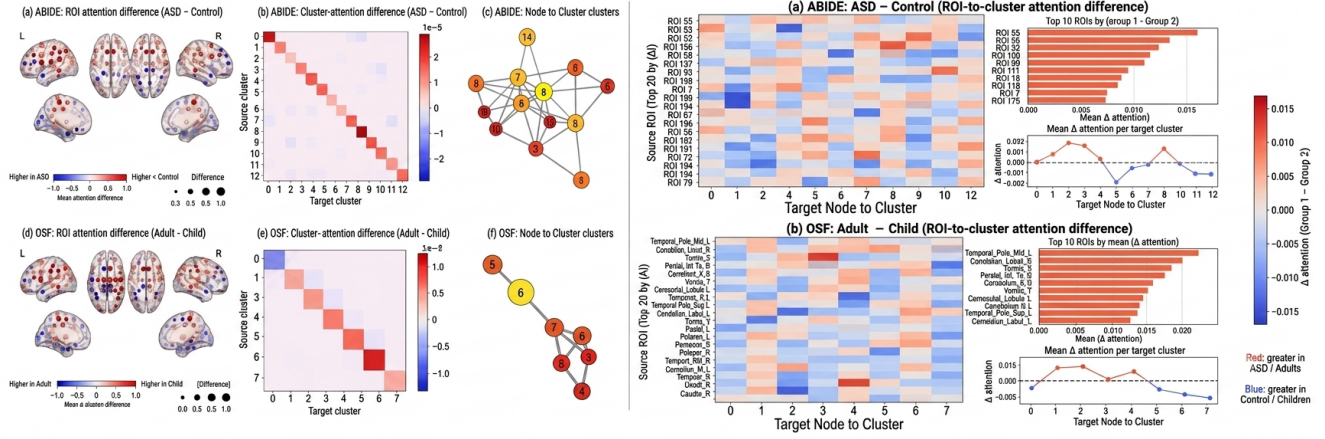


Figure 3: **Analysis of group-specific Multiscale attention.** The **left block (a–f)** shows ROI-level attention differences, source-cluster-aggregated ROI-to-target-cluster differences, and the corresponding FC clustering topology for ABIDE and OSF. The **right block (a–b)** provides the detailed ROI-to-cluster analysis, including the 20 ROIs with the largest mean absolute differences, the 10 most differing ROIs, and the mean signed attention difference across target Node to cluster modules. Differences are computed as ASD–Control for ABIDE and Adult–Child for OSF. **Red** denotes greater attention in ASD or adults, whereas **Blue** denotes greater attention in controls or children.

Table 2: Alternate-atlas evaluation on ABIDE and OSF. Best results for each dataset are shown in bold.

Dataset	Method	ACC(%) $\uparrow$	AUC(%) $\uparrow$	SEN(%) $\uparrow$	SPE(%) $\uparrow$
ABIDE (AAL)	ALTER	61.07	67.55	63.24	58.73
	BrainRMA-S	61.07	67.04	64.71	57.14
	BrainRMA-L	61.83	68.51	57.35	66.67
OSF (Craddock)	ALTER	87.50	88.00	85.71	88.00
	BrainRMA-S	84.38	89.14	71.43	88.00
	BrainRMA-L	90.63	87.43	71.43	96.00

subjects. ROI partitions were then generated using the corresponding clustering strategy and fixed for validation and testing. Where applicable, the number of clusters was determined as $K = \lceil N/p \rceil$, where N is the total number of ROIs and p is the target patch size. The model uses four attention heads and was trained for 200 epochs using Adam with a cosine learning-rate scheduler, a learning rate and weight decay of 10^{-4} , and batch size 64. We release source code (See Supplementary) for complementing the implementation details.

Evaluation Metrics We report accuracy (ACC), area under the ROC curve (AUC), sensitivity (SEN) or true positive rate, and specificity (SPE) or true negative rate.

5 Results and Discussion

Comparison with SOTA Baselines

Table 1 compares BrainRMA with recent learnable-network and graph-transformer baselines. On ABIDE, BrainRMA with Linear Kernel Attention (BrainRMA-L) achieves the best results across all metrics, reaching 82.35% ACC and 85.36% AUC. On OSF, both variants attain 93.75% ACC, while BrainRMA-L achieves the highest AUC of 90.86% and specificity of 96.00%. Compared with the strongest baseline,

BrainRMA-L improves ACC by 3.35 percentage points on ABIDE and 6.25 points on OSF, demonstrating the effectiveness of jointly modeling global ROI dependencies and ROI-cluster interactions, and achieving a new SOTA.

Multiscale ROI-Cluster Interpretability

BrainRMA provides interpretation at 3 complementary scales: ROI-level attention, ROI-to-cluster interactions, and cluster-level organization. Group-wise attention differences are computed as ASD–Control for ABIDE and Adult–Child for OSF. In Figure 3, the left block (a)–(f) presents ROI-level differences, inter-cluster attention, and functional-module topology, while the right block (a)–(b) highlights the top 20 most discriminative ROIs, the top 10 ranked regions, and their signed attention differences across Node to cluster attentions. For ABIDE, the most discriminative ROIs identified by BrainRMA span temporal, frontal, parietal, visual-sensory, and default-mode systems which are associated with social cognition, sensory integration, and theory-of-mind processing as reported in neurobiology literature (Di Martino et al. 2009; Shih et al. 2011; Koldewyn, Whitney, and Rivera 2011; Marco et al. 2011; Padmanabhan et al. 2017; Wang et al. 2021). For OSF, the prominent attention differences identified by BrainRMA are in the temporal, inferior frontal, paracentral, cerebellar, and vermal regions which are linked to developmental changes from child to adult in social reasoning and cognitive control (Rubia et al. 2007; Stoodley and Schmahmann 2010; Ross and Olson 2010; Richardson et al. 2018). Overall, the analysis shows that BrainRMA predictions arise from distributed ROI-cluster interactions rather than isolated regions, and are well aligned with neurological findings. Note that these patterns are model-derived associations, not causal biomarkers.

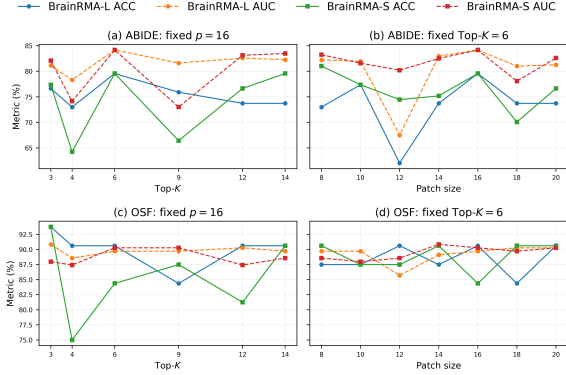


Figure 4: **Effect of patch size and Top- K on ACC and AUC.** Panels (a,c) vary Top- K with $p = 16$, while (b,d) vary patch size with Top- $K = 6$. Panels (a,b) show ABIDE and (c,d) show OSF. Solid/dashed lines denote ACC/AUC, and circles/squares denote BrainRMA-L/BrainRMA-S.

Ablation Studies

Benefit of Multiscale The long-range-only (global) model achieves **75.18%** and **90.63%** accuracy on ABIDE and OSF, respectively, while the cluster-aware-only (local) model achieves **63.50%** and **81.25%**, showing the benefit of multiscale attention.

Atlas Selection To assess robustness to brain parcellation, we evaluate the proposed variants using the AAL atlas for ABIDE and the Craddock atlas for OSF. Since the implementation of BrainCKT is not publicly available, we use ALTER as the closest reproducible and competitive baseline. As shown in Table 2, both BrainRMA variants remain competitive under the alternate-atlas settings. BrainRMA-L performs best in overall accuracy, whereas BrainRMA-S provides stronger sensitivity on ABIDE and AUC on OSF, suggesting that the preferred attention formulation may depend on the dataset and atlas granularity.

Patch Size and Top- K We study the effect of functional-module granularity and group-graph sparsity by varying the patch size over $\{8, 10, 12, 14, 16, 18, 20\}$ and Top- K over $\{3, 4, 6, 9, 12, 14\}$. Patch size controls the approximate number of ROIs per module, whereas Top- K determines the sparsity of the training-derived functional-connectivity graph.

Figure 4 summarizes the ACC and AUC trends by varying one parameter while fixing the other. On ABIDE, BrainRMA reaches its strongest AUC at (20, 3), while BrainRMA-S performs best around (8, 6). On OSF, both variants remain comparatively stable, with their strongest configurations obtained near (16, 3) for BrainRMA-L and (8, 9) for BrainRMA-S. Overall, the preferred partition depends on both the dataset and attention formulation, with ABIDE showing greater sensitivity than OSF. Complete grid results, including sensitivity and specificity, are provided in supplementary.

Clustering Method We evaluate ROI partitioning using METIS (Karypis and Kumar 1998), K-means (Lloyd 1982), spectral clustering (Ng, Jordan, and Weiss 2002), and Lou-

Table 3: Clustering ablation on ABIDE and OSF. METIS, K-m., Spec., Louv., Anat., Yeo, DMoN, and MinCut denote METIS, K-means, spectral clustering, Louvain clustering, anatomical grouping, Yeo-network grouping, DMoN, and MinCut pooling, respectively. S and L denote BrainRMA-S and BrainRMA-L. Bold indicates the best result for each dataset, attention variant, and metric.

Dataset	Attn.	Metric	Heuristic			Biological		Learnable	
			METIS	K-m.	Spec.	Louv.	Anat.	Yeo	DMoN MinCut
ABIDE	S	ACC	67.15	74.45	67.88	67.88	73.72	72.26	67.88
		AUC	76.65	81.86	81.24	79.64	81.94	79.70	78.98
		SEN	44.29	62.86	47.14	85.71	68.57	80.00	71.43
		SPE	91.04	86.57	89.55	49.25	79.10	64.18	73.13
	L	ACC	82.35	75.91	73.72	67.88	71.53	67.15	67.88
		AUC	85.36	83.20	83.71	79.72	79.74	78.57	79.94
		SEN	82.61	70.00	62.86	50.00	67.14	52.86	47.14
		SPE	82.09	82.09	85.07	86.57	76.12	82.09	89.55
OSF	S	ACC	93.75	90.63	90.63	87.50	93.75	87.50	90.63
		AUC	88.00	90.86	89.14	88.57	89.14	88.57	90.86
		SEN	85.71	85.71	85.71	71.43	85.71	71.43	85.71
		SPE	96.00	92.00	92.00	92.00	96.00	92.00	92.00
	L	ACC	84.37	84.38	78.12	90.62	71.88	87.50	87.50
		AUC	84.57	89.71	86.86	90.29	82.86	90.29	92.00
		SEN	71.42	71.43	85.71	85.71	85.71	71.43	85.71
		SPE	88.00	88.00	76.00	92.00	68.00	92.00	88.00

vain community detection (Blondel et al. 2008). We also compare atlas-based anatomical grouping (Tzourio-Mazoyer et al. 2002), Yeo functional-network grouping (Yeo et al. 2011), and the learnable DMoN (Tsitsulin et al. 2023) and MinCut (Bianchi, Grattarola, and Alippi 2020) methods. The remaining architecture is fixed, and all data-driven partitions are constructed using only the training subjects. Table 3 shows that both BrainRMA variants are affected by the ROI partition. On ABIDE, METIS is strongest for BrainRMA-L, whereas K-means and anatomical grouping perform best for BrainRMA-S in accuracy and AUC, respectively. On OSF, several strategies remain competitive: METIS and anatomical grouping are strongest for BrainRMA-S, while MinCut provides the best AUC for BrainRMA-L. Overall, ABIDE is more sensitive to partition choice, whereas OSF is more robust across clustering strategies.

6 Conclusion

We introduce BrainRMA, an ROI-preserving framework that jointly models long-range ROI interactions, ROI-to-cluster communication, and cluster-level organization to provide a comprehensive view of the brain communication. By restoring cluster-updated features to their original ROI indices, BrainRMA combines these complementary views without losing region-specific information, supporting both prediction and multiscale interpretation. Future work will examine larger multi-site and longitudinal datasets, additional brain atlases, multimodal imaging, and stronger neurobiological validation of the learned attention patterns.

References

- Bannadabhavi, A.; Lee, S.; Deng, W.; Ying, R.; and Li, X. 2023. Community-Aware Transformer for Autism Prediction in fMRI Connectome. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 287–297. Springer.
- Bertolero, M. A.; Yeo, B. T.; and D’Esposito, M. 2015. The modular and integrative functional architecture of the human brain. *Proceedings of the National Academy of Sciences*, 112(49): E6798–E6807.
- Bianchi, F. M.; Grattarola, D.; and Alippi, C. 2020. Spectral Clustering with Graph Neural Networks for Graph Pooling. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, 874–883. PMLR.
- Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.
- Craddock, R. C.; Benhajali, Y.; Chu, C.; Chouinard, F.; Evans, A.; Jakab, A.; Khundrakpam, B. S.; Lewis, J. D.; Li, Q.; Milham, M.; Yan, C.; and Bellec, P. 2013. The Neuro Bureau Preprocessing Initiative: Open Sharing of Preprocessed Neuroimaging Data and Derivatives. In *Neuroinformatics 2013*. Stockholm, Sweden.
- Cui, H.; Dai, W.; Zhu, Y.; Kan, X.; Gu, A. A. C.; Lukemire, J.; Zhan, L.; He, L.; Guo, Y.; and Yang, C. 2022. BrainGB: A Benchmark for Brain Network Analysis with Graph Neural Networks. *IEEE Transactions on Medical Imaging*, 42(2): 493–506.
- Di Martino, A.; Ross, K.; Uddin, L. Q.; Sklar, A. B.; Castellanos, F. X.; and Milham, M. P. 2009. Functional brain correlates of social and nonsocial processes in autism spectrum disorders: an activation likelihood estimation meta-analysis. *Biological Psychiatry*, 65(1): 63–74.
- Di Martino, A.; Yan, C.-G.; Li, Q.; Denio, E.; Castellanos, F. X.; Alaerts, K.; Anderson, J. S.; Assaf, M.; Bookheimer, S. Y.; Dapretto, M.; Deen, B.; Delmonte, S.; Dinstein, I.; Ertl-Wagner, B.; Fair, D. A.; Gallagher, L.; Kennedy, D. P.; Keown, C. L.; Keyzers, C.; Lainhart, J. E.; Lord, C.; Luna, B.; Menon, V.; Minshew, N. J.; Monk, C. S.; Mueller, S.; Müller, R.-A.; Nebel, M. B.; Nigg, J. T.; O’Hearn, K.; Pelphrey, K. A.; Peltier, S. J.; Rudie, J. D.; Sunaert, S.; Thioux, M.; Tyszka, J. M.; Uddin, L. Q.; Verhoeven, J. S.; Wenderoth, N.; Wiggins, J. L.; Mostofsky, S. H.; and Milham, M. P. 2014. The Autism Brain Imaging Data Exchange: Towards a Large-Scale Evaluation of the Intrinsic Brain Architecture in Autism. *Molecular Psychiatry*, 19(6): 659–667.
- Dwivedi, V. P.; and Bresson, X. 2020. A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*.
- Fair, D. A.; Cohen, A. L.; Power, J. D.; Dosenbach, N. U. F.; Church, J. A.; Miezin, F. M.; Schlaggar, B. L.; and Petersen, S. E. 2009. Functional Brain Networks Develop from a “Local to Distributed” Organization. *PLOS Computational Biology*, 5(5): e1000381.
- Gallos, L. K.; Makse, H. A.; and Sigman, M. 2012. A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proceedings of the National Academy of Sciences*, 109(8): 2825–2830.
- Hamilton. 2017. Inductive Representation Learning on Large Graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 1024–1034.
- Huang, S.; Song, Y.; Zhou, J.; and Lin, Z. 2024. Cluster-wise Graph Transformer with Dual-granularity Kernelized Attention. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- Kan, X.; Cui, H.; Lukemire, J.; Guo, Y.; and Yang, C. 2022a. FB-NETGEN: Task-Aware GNN-Based fMRI Analysis via Functional Brain Network Generation. In *International Conference on Medical Imaging with Deep Learning (MIDL)*, 618–637.
- Kan, X.; Dai, W.; Cui, H.; Zhang, Z.; Guo, Y.; and Yang, C. 2022b. Brain Network Transformer. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, 25586–25599.
- Karypis, G.; and Kumar, V. 1998. A Fast and High Quality Multi-level Scheme for Partitioning Irregular Graphs. *SIAM Journal on Scientific Computing*, 20(1): 359–392.
- Kawahara, J.; Brown, C. J.; Miller, S. P.; Booth, B. G.; Chau, V.; Grunau, R. E.; Zwicker, J. G.; and Hamarneh, G. 2017. Brain-NetCNN: Convolutional Neural Networks for Brain Networks; Towards Predicting Neurodevelopment. *NeuroImage*, 146: 1038–1049.
- Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.
- Koldewyn, K.; Whitney, D.; and Rivera, S. M. 2011. Neural correlates of coherent and biological motion perception in autism. *Developmental Science*, 14(5): 1075–1088.
- Kreuzer, D.; Beaini, D.; Hamilton, W.; Létourneau, V.; and Tossou, P. 2021. Rethinking graph transformers with spectral attention. *Advances in neural information processing systems*, 34: 21618–21629.
- Li, X.; Zhou, Y.; Dvornek, N.; Zhang, M.; Gao, S.; Zhuang, J.; Scheinost, D.; Staib, L. H.; Ventola, P.; and Duncan, J. S. 2021. BrainGNN: Interpretable Brain Graph Neural Network for fMRI Analysis. *Medical Image Analysis*, 74: 102233.
- Lloyd, S. P. 1982. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2): 129–137.
- Marco, E. J.; Hinkley, L. B.; Hill, S. S.; and Nagarajan, S. S. 2011. Sensory processing in autism: a review of neurophysiologic findings. *Pediatric Research*, 69(5 Pt 2): 48R–54R.
- Meng, L.; Liang, K.; Yu, H.; Wang, H.; Li, M.; and Liu, X. 2026. Make Model Transparent: Brain Network Analysis via Causal and Knowledge Graph Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, 24362–24369.
- Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2002. On Spectral Clustering: Analysis and an Algorithm. In *Advances in Neural Information Processing Systems*, volume 14, 849–856.
- Padmanabhan, A.; Lynch, C. J.; Schaer, M.; and Menon, V. 2017. The default mode network in autism. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 2(6): 476–486.
- Peng, Z.; He, Z.; Jiang, Y.; Wang, P.; and Yuan, Y. 2024. GBT: Geometric-Oriented Brain Transformer for Autism Diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 142–152. Springer.
- Power, J. D.; Cohen, A. L.; Nelson, S. M.; Wig, G. S.; Barnes, K. A.; Church, J. A.; Vogel, A. C.; Laumann, T. O.; Miezin, F. M.; Schlaggar, B. L.; et al. 2011. Functional network organization of the human brain. *Neuron*, 72(4): 665–678.
- Richardson, H.; Lisandrelli, G.; Riobueno-Naylor, A.; and Saxe, R. 2018. Development of the Social Brain from Age Three to Twelve Years. *Nature Communications*, 9(1): 1027.
- Ross, L. A.; and Olson, I. R. 2010. Social Cognition and the Anterior Temporal Lobes. *NeuroImage*, 49(4): 3452–3462.
- Rubia, K.; Smith, A. B.; Taylor, E.; and Brammer, M. 2007. Linear Age-Related Functional Development of Right Inferior Frontostriato-Cerebellar Networks during Response Inhibition and Anterior Cingulate during Error-Related Processes. *Human Brain Mapping*, 28(11): 1163–1177.

Rubinov, M.; and Sporns, O. 2010. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3): 1059–1069.

Rudie, J. D.; Brown, J.; Beck-Pancer, D.; Hernandez, L.; Dennis, E.; Thompson, P.; Bookheimer, S.; and Dapretto, M. 2013. Altered functional and structural brain network organization in autism. *NeuroImage: clinical*, 2: 79–94.

Shih, P.; Keehn, B.; Oram, J. K.; Leyden, K. M.; Keown, C. L.; and Müller, R.-A. 2011. Functional differentiation of posterior superior temporal sulcus in autism: a functional connectivity magnetic resonance imaging study. *Biological Psychiatry*, 70(3): 270–277.

Stoodley, C. J.; and Schmahmann, J. D. 2010. Evidence for Topographic Organization in the Cerebellum of Motor Control versus Cognitive and Affective Processing. *Cortex*, 46(7): 831–844.

Tsitsulin, A.; Palowitch, J.; Perozzi, B.; and Müller, E. 2023. Graph Clustering with Graph Neural Networks. *Journal of Machine Learning Research*, 24(127): 1–21.

Tzourio-Mazoyer, N.; Landeau, B.; Papathanassiou, D.; Crivello, F.; Etard, O.; Delcroix, N.; Mazoyer, B.; and Joliot, M. 2002. Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, 15(1): 273–289.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention Is All You Need. [arXiv:1706.03762](https://arxiv.org/abs/1706.03762).

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph Attention Networks. In *International Conference on Learning Representations (ICLR)*.

Wang, M.; et al. 2021. Resting-state abnormalities in functional connectivity of the default mode network in autism spectrum disorder: a meta-analysis. *Brain Imaging and Behavior*.

Wu, Z.; Jain, P.; Wright, M.; Mirhoseini, A.; Gonzalez, J. E.; and Stoica, I. 2021. Representing long-range context for graph neural networks with global attention. *Advances in neural information processing systems*, 34: 13266–13279.

Xie, J.; Girshick, R.; and Farhadi, A. 2016. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, 478–487. PMLR.

Yang, B.; Wang, M.; Zhou, W.; Wang, X.; Chen, S.; Potenza, M. N.; Yuan, L.-X.; and Dong, G.-H. 2023. Disrupted network integration and segregation involving the default mode network in autism spectrum disorder. *Journal of Affective Disorders*, 323: 309–319.

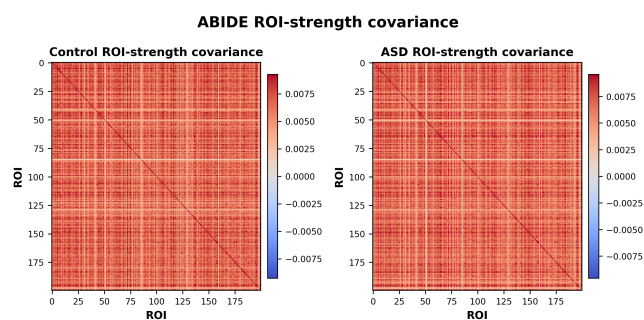
Yang, Y.; Zhao, B.; Ni, Z.; Zhao, Y.; and Li, X. 2024. Learnable Community-Aware Transformer for Brain Connectome Analysis with Token Clustering. *arXiv preprint arXiv:2403.08203*.

Yeo, B. T. T.; Krienen, F. M.; Sepulcre, J.; Sabuncu, M. R.; Lashkari, D.; Hollinshead, M.; Roffman, J. L.; Smoller, J. W.; Zöllei, L.; Polimeni, J. R.; Fischl, B.; Liu, H.; and Buckner, R. L. 2011. The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity. *Journal of Neurophysiology*, 106(3): 1125–1165.

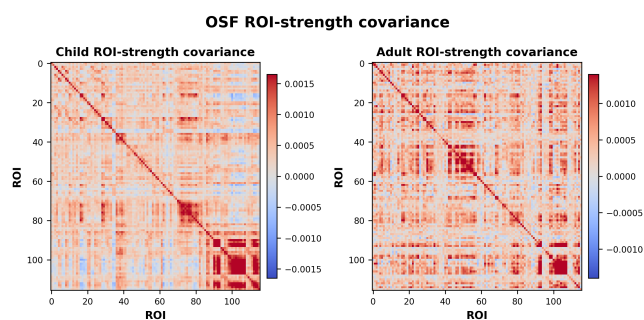
Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; and Liu, T.-Y. 2021. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*, 34: 28877–28888.

Yu, S.; Jin, S.; Li, M.; Sarwar, T.; and Xia, F. 2024. Long-Range Brain Graph Transformer. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, 24472–24495.

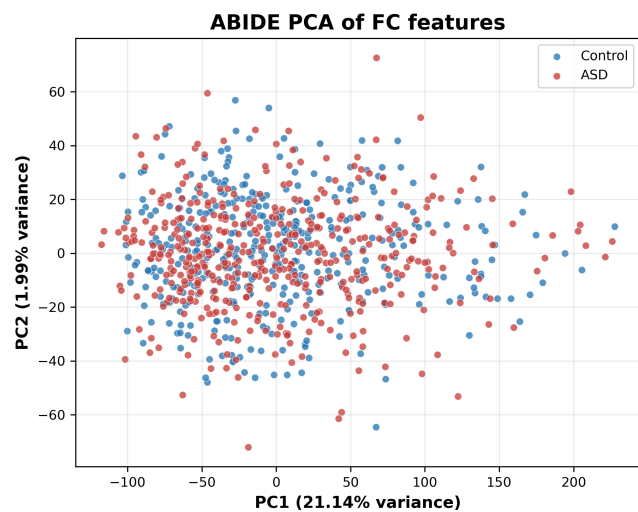
A Supplementary Material



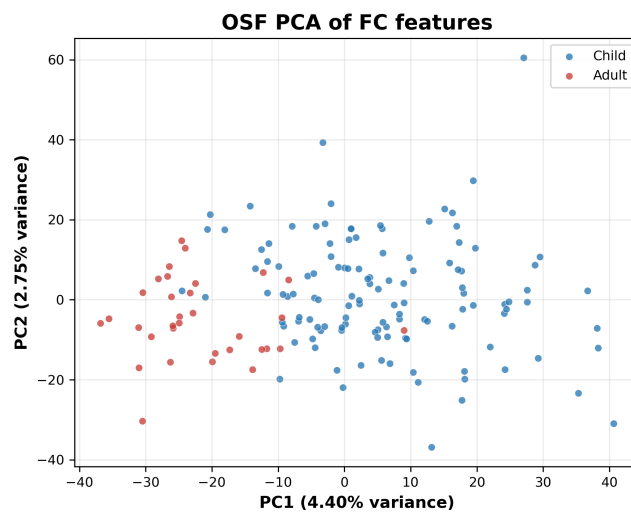
(a) ABIDE covariance



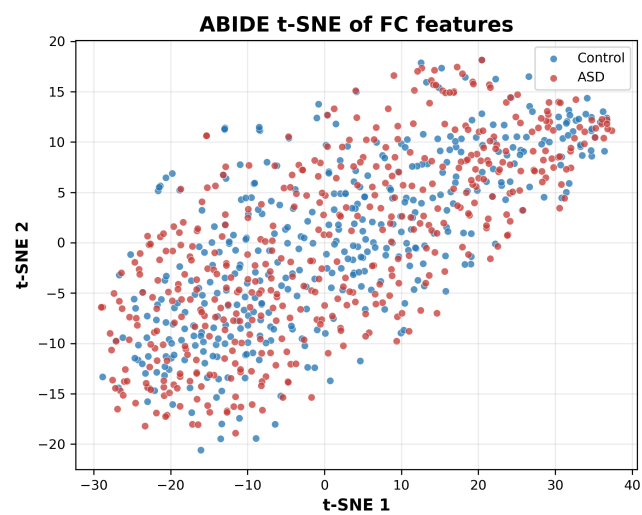
(b) OSF covariance



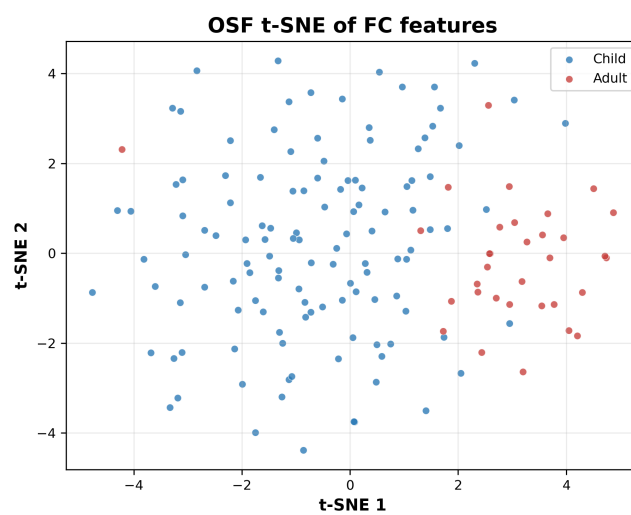
(c) ABIDE PCA



(d) OSF PCA



(e) ABIDE t-SNE



(f) OSF t-SNE

Figure S1: Exploratory analysis of functional-connectivity features on ABIDE and OSF. Panels (a) and (b) show group-specific ROI-strength covariance patterns. Panels (c) and (d) show PCA projections, and panels (e) and (f) show t-SNE projections of the functional-connectivity features. ABIDE shows stronger ASD–Control overlap, indicating a more challenging disease-classification setting, whereas OSF shows clearer Adult–Child separation, suggesting stronger developmental structure in the feature space.

<

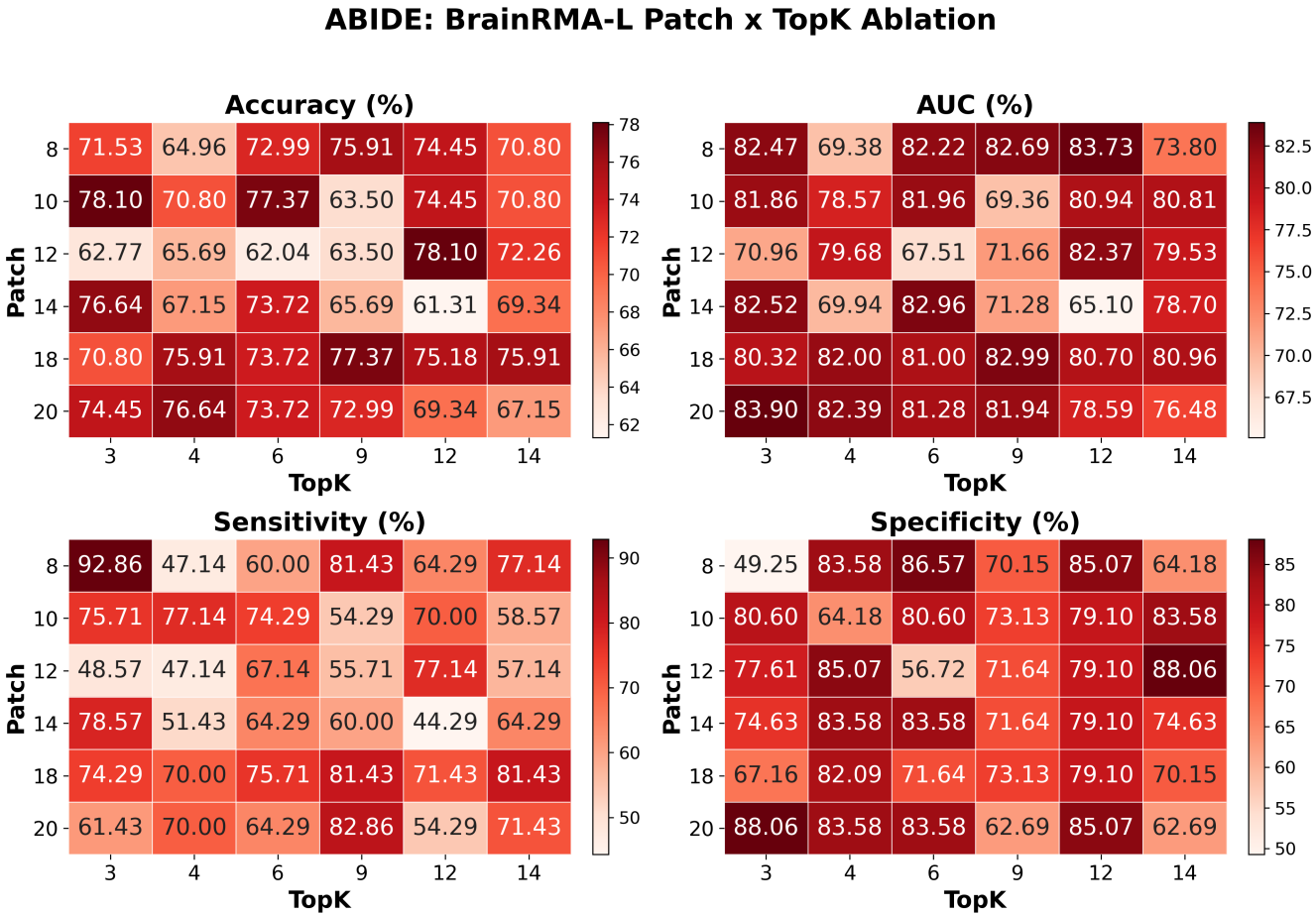


Figure S2: Patch-size and Top-*K* ablation analysis for BrainRMA-L on ABIDE. Each heatmap reports accuracy, AUC, sensitivity, and specificity across different functional-cluster granularities and graph-sparsity levels. Patch size controls the target average number of ROIs assigned to each functional cluster, while Top-*K* controls the number of strongest functional-connectivity neighbours retained per ROI before clustering. The results show that performance is sensitive to both cluster granularity and graph sparsity.

<

OSF: BrainRMA-L Patch x TopK Ablation

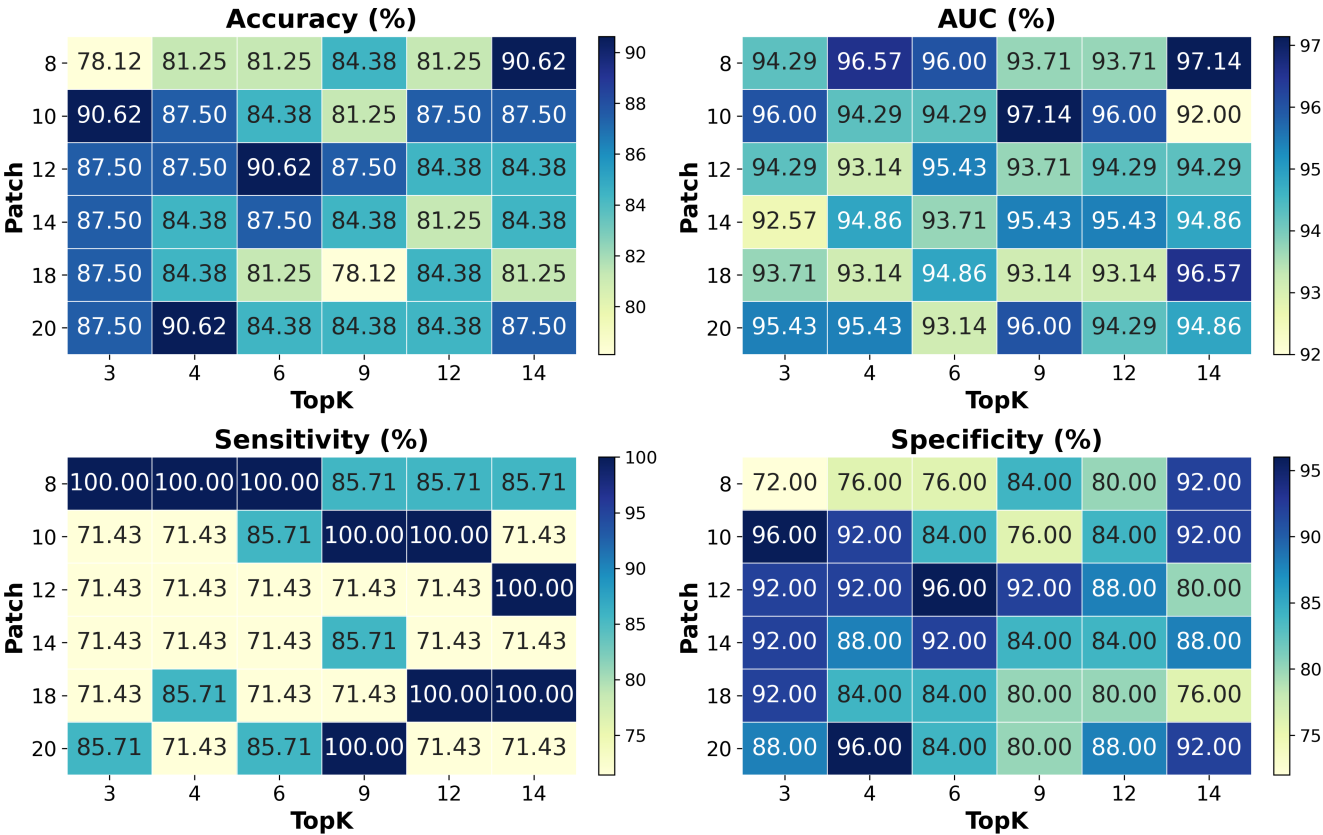


Figure S3: Patch-size and Top- K ablation analysis for BrainRMA-L on OSF. The heatmaps report accuracy, AUC, sensitivity, and specificity across different functional-cluster granularities and graph-sparsity levels. Strong performance across several settings suggests that the linear node-to-cluster variant captures stable developmental functional-connectivity patterns when the clusters provide meaningful module-level context.

<

ABIDE: BrainRMA-S Patch x TopK Ablation

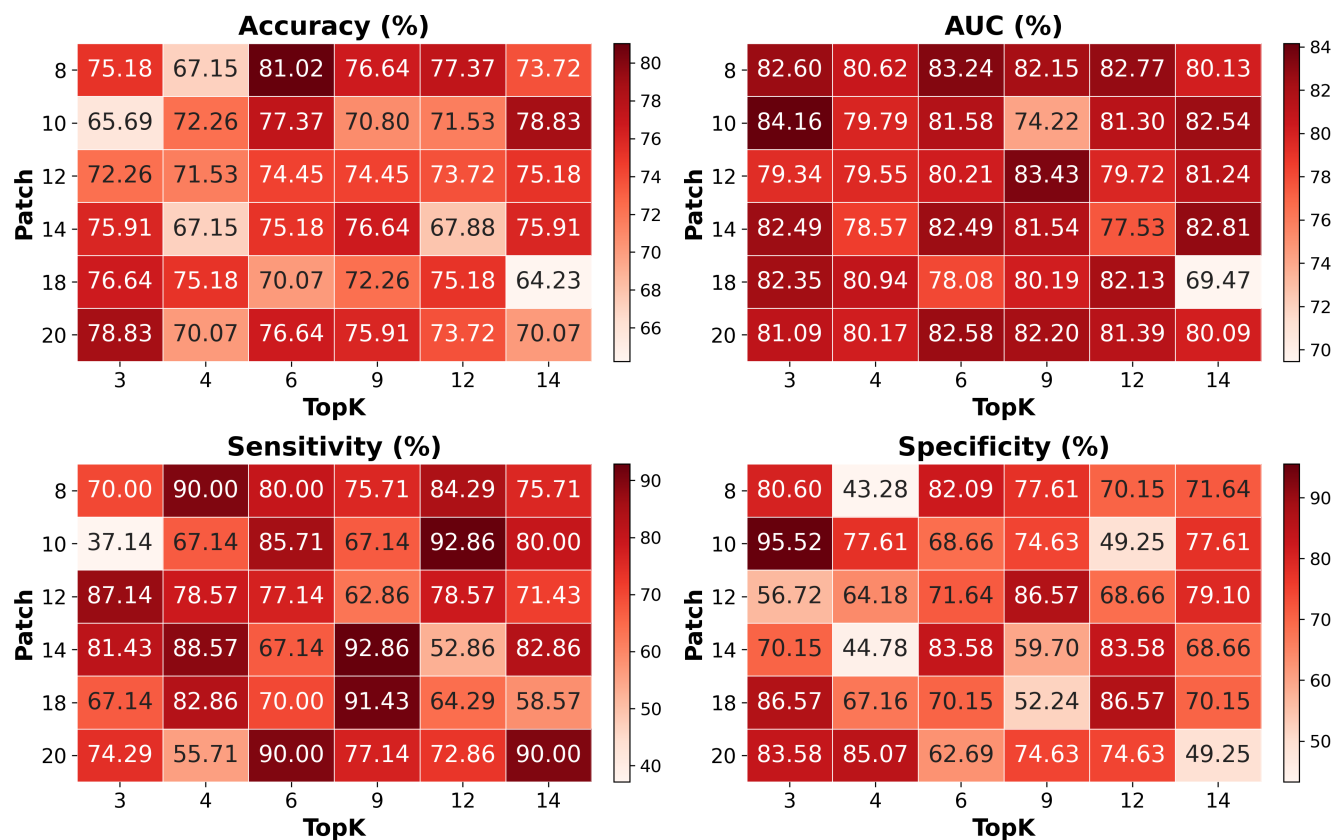


Figure S4: Patch-size and Top- K ablation analysis for BrainRMA-S on ABIDE. BrainRMA-S employs softmax-based dot-product node-to-cluster attention. The heatmaps report accuracy, AUC, sensitivity, and specificity across different patch-size and Top- K combinations. The results highlight the importance of selecting cluster settings that preserve disease-relevant ROI-level and module-level information.

OSF: BrainRMA-S Patch x TopK Ablation

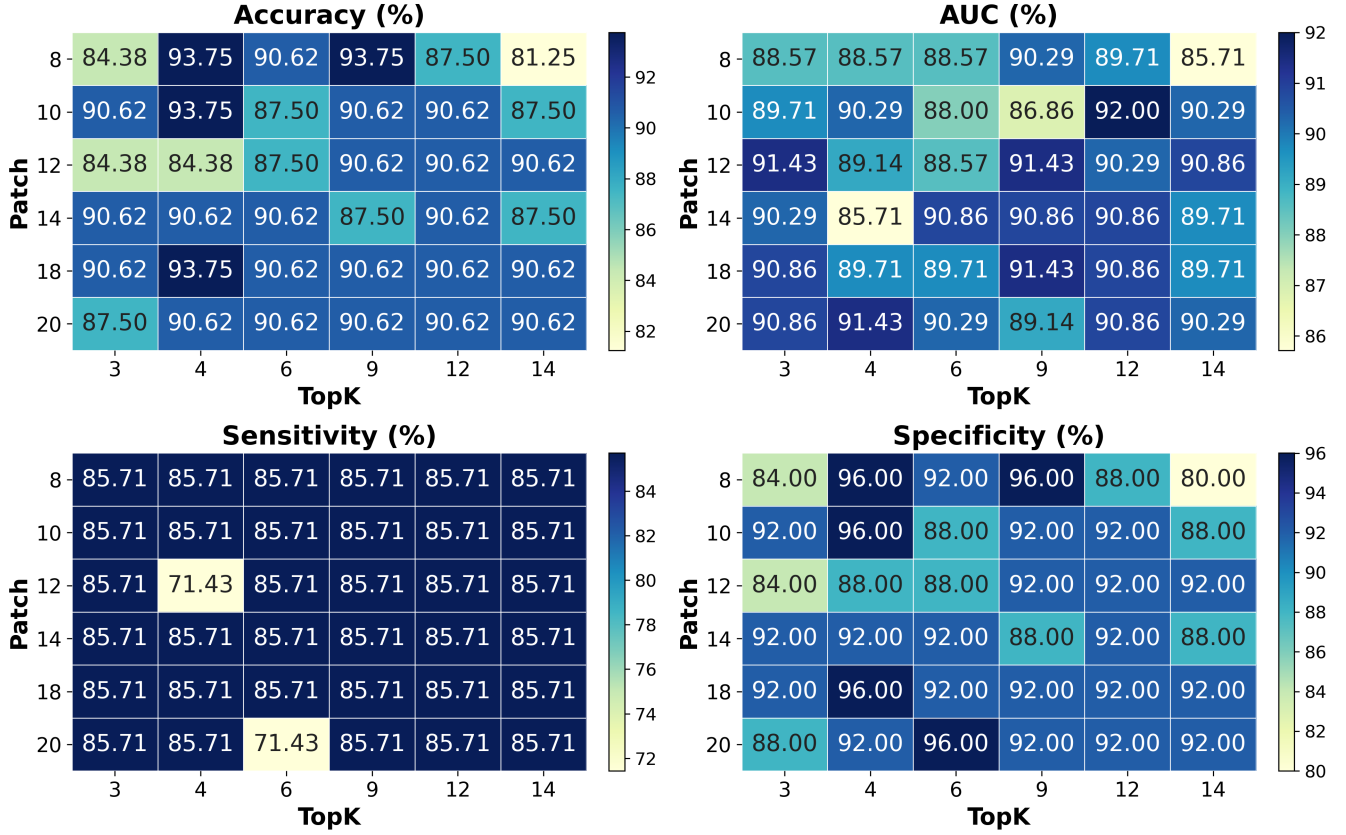


Figure S5: Patch-size and Top- K ablation analysis for BrainRMA-S on OSF. The heatmaps summarize accuracy, AUC, sensitivity, and specificity across different functional-cluster granularities and graph-sparsity levels. Compared with ABIDE, OSF shows more stable performance across many patch-size and Top- K combinations, suggesting more separable developmental functional-connectivity differences.

Additional Experimental Analysis

Figure S1 provides an exploratory comparison of the functional-connectivity feature spaces for ABIDE and OSF. The ROI-strength covariance maps show that ABIDE has highly similar global covariance structures for ASD and control subjects, suggesting that the diagnostic differences are subtle and distributed across multiple ROI interactions. This is also reflected in the PCA and t-SNE projections, where ASD and control samples substantially overlap. ABIDE therefore represents a more challenging classification setting in which the model must learn weak but coordinated connectivity patterns across multiple brain systems.

In contrast, OSF shows clearer differences between child and adult subjects. Its ROI-strength covariance maps reveal more visible group-level variation, while the PCA and t-SNE projections show stronger separation between the two age groups. This suggests that developmental differences are more strongly expressed in the OSF functional-connectivity feature space than the ASD–Control differences in ABIDE. These observations help explain the comparatively stronger

classification performance obtained on OSF.

Figures S2–S5 present the complete patch-size and Top- K ablation results for BrainRMA-L and BrainRMA-S on ABIDE and OSF. Patch size determines the target average number of ROIs in each functional cluster, whereas Top- K determines the sparsity of the group-level functional graph used for clustering. Moderate settings generally provide stronger performance, suggesting that overly small clusters may fragment functional systems, while overly large clusters may merge distinct ROI-level information. These findings support preserving ROI-level representations while incorporating cluster-level functional context.

B Implementation Details

The implementation is provided in `code.zip`. The released code implements BrainRMA for functional brain graph classification on ABIDE and OSF. Each subject is represented as a graph whose nodes are atlas-defined ROIs, and each ROI feature is derived from its whole-brain functional-connectivity profile. The code constructs training-derived

FC-METIS clusters, applies the long-range ROI interaction branch and the ROI-preserving node-to-cluster branch, fuses the two ROI-aligned representations, and performs subject-level classification using the graph readout module.

The main hyperparameters used in the experiments are the patch size p , Top- K sparsification value k_{top} , hidden dimension, number of attention heads, dropout rate, batch size, learning rate, weight decay, number of epochs, and residual gate initialization. In the reported setting, the model uses FC-METIS clustering with patch size $p = 16$ and Top- $K = 8$. The model is trained using the Adam optimizer with cross-entropy loss for supervised classification. The provided scripts also support running BrainRMA-L and BrainRMA-S separately, where BrainRMA-L uses the linear node-to-cluster attention variant and BrainRMA-S uses the softmax/dot-product node-to-cluster attention variant.

All dataset splits are loaded from fixed train/validation/test split files. For each split, the group-level functional-connectivity graph and ROI clusters are constructed only from the training subjects and then kept fixed for validation and testing. This avoids using validation or test information during cluster construction.