HADBN: LEARNABLE HIERARCHICAL AGGREGATED DYNAMIC BRAIN NETWORKS FOR INTERPRETABLE BRAIN DISORDER DIAGNOSIS

Anonymous authors
Paper under double-blind review

Keywords:
Spatio-temporal Graph neural networks,
Hierarchical aggregation,
Dynamic brain networks,
Brain disorder diagnosis

ABSTRACT

The complex spatio-temporal correlations in brain networks make brain disorder diagnosis challenging. Many spatio-temporal models underutilize anatomical constraints from brain spatial topology, limiting their ability to capture spatial relationships and dynamic interactions. To address this, we propose a novel framework, Learnable Hierarchical Aggregated Dynamic Brain Networks (HADBN), which incorporates multi-level anatomical priors of brain spatial topology and jointly learns multi-scale spatial representations and time-varying connectivity patterns. Specifically, HADBN constructs spatio-temporal interaction graph neural networks (GNNs) with a three-level hierarchical brain architecture—local ROIs, middle sub-network, and global network levels—guided by functional parcellation atlases. Unlike existing multi-level spatio-temporal models, HADBN introduces learnable F-matrix aggregation modules that perform data-driven pooling across spatial scales, replacing fixed anatomical aggregation. Moreover, it employs a feature-driven temporal attention mechanism that computes adaptive window weights from multi-scale features, thereby enabling principled spatio-temporal integration and cross-level information flow. Experiments show that HADBN consistently outperforms state-of-the-art methods under 10-fold cross-validation, while the learned aggregation weights offer interpretable insights into disorder-relevant brain networks and support biomarker discovery. These results establish HADBN as a robust and interpretable framework for brain disorder diagnosis.

1 INTRODUCTION

The human brain is a complex network characterized by hierarchical organization and real-time dynamic interactions, making brain disorder diagnosis challenging. Resting-state functional magnetic resonance imaging (rs-fMRI) has emerged as a non-invasive technique for probing the neural correlates of brain disorders. It captures spontaneous fluctuations in blood-oxygen-level-dependent (BOLD) signals that reflect intrinsic brain activity(Biswal & Uddin, 2025). From BOLD signals, functional connectivity networks (FCNs) are constructed to quantify statistical dependencies between brain regions(Du et al., 2024).

FCNs include static functional connectivity (sFC) and dynamic FC (dFC). sFC estimates correlations between brain regions by averaging BOLD signals over the time series (Xie et al., 2026), implicitly treating functional interactions as stationary. To capture time-varying interactions between brain regions, dFC computes correlations over short time windows (Laumann et al., 2024). Thus, dFC reveals how the strength and topology of functional connections fluctuate over time and has shown sensitivity to disorder-specific patterns (Wen et al., 2026). Thus, capturing temporal variability is essential for developing robust and sensitive diagnostic models.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Given the topological organization of the brain, an FCN can be represented as a graph in which nodes denote brain regions and edges denote functional connections between them(Tang et al., 2026). Consequently, graph neural networks (GNNs) have gained significant attention for graph representation learning, leveraging the inherent graph structure to capture both node-level features and topological relationships simultaneously(Li et al., 2025a). To further capture spatiotemporal information, recent studies have extended GNNs to spatiotemporal GNNs (STGNNs) that process sequences of dFC matrices(Liu et al., 2024c). These STGNNs typically rely on classical GNN architectures for spatial modeling and use recurrent neural networks, temporal convolution layers, or self-attention mechanisms for temporal information learning (Liu et al., 2025b).

However, a key challenge for STGNNs is how to adequately represent the spatiotemporal information of brain networks. Brain information processing follows a hierarchical pattern involving local and long-range interactions between brain regions, which is organized from local regions to the whole-brain network. Different hierarchical brain regions have distinct functions and specific temporal patterns, which gives rise to complex cognitive and behavioral functions through hierarchical dynamic coordination (Zhou et al., 2026; Ma et al., 2026a). Therefore, spatiotemporal representation learning based on hierarchical brain networks is promising for brain disorder diagnosis.

The construction of hierarchical brain networks can be broadly divided into non-prior-guided and prior-guided approaches. Non-prior-guided hierarchical brain networks are learned by the model itself according to the learning objective (e.g., classification accuracy), without prior guidance from medical brain parcellation. For example, Zhou et al. (2026) proposed a multi-scale spatiotemporal fusion neural network (MS-STFNN) to represent multi-granularity spatial hierarchical features of the brain from local to global levels for major depressive disorder diagnosis. The multi-granularity spatial hierarchical features include fine-grained node connections, coarse-grained module interactions, and intra-modular topological features, where the number of coarse-grained modules is pre-specified and determined by grid search. Although such spatial hierarchies are learnable during model training, the lack of prior guidance limits their interpretability for brain disorder diagnosis.

For prior-guided hierarchical brain networks, brain regions are divided into different levels based on prior brain atlases. Ma et al. (2026) used prior-guided brain region clustering for autism spectrum disorder (ASD) and Alzheimer's disease (AD) diagnosis, where the lower level of the prior brain hierarchy was defined by the Automated Anatomical Labeling (AAL) atlas and the upper level by the Yeo-7 network atlas. Similarly, MHNet (Li et al., 2025c) divided the brain network into local, intermediate, and global levels based on AAL and Yeo-7 brain atlases, and constructed functional brain graphs at each level using sFC, thereby enhancing the interpretability of the diagnostic model. Although these prior-guided hierarchical brain network studies achieved good results in brain disorder classification, they have not fully considered the temporal information embedded in dFC and thus have not simultaneously learned spatiotemporal representations of brain networks.

To address this, we propose a spatiotemporal representation learning framework, called Learnable Hierarchical Aggregated Dynamic Brain Networks (HADBN), based on prior-guided hierarchical brain networks for brain disorder diagnosis. Our main contributions are as follows:

(a) We design a multi-view brain network to construct hierarchical graph representations at three progressive levels: the local region-of-interest (ROI) level (LL), the middle sub-network level (ML), and the global large-scale network level (GL). Each hierarchical graph is guided by prior functional parcellation atlases. To aggregate nodes (brain regions) from lower- to higher-level graphs, we design learnable F-matrices for ML and GL to replace fixed anatomical aggregation. These F-matrices are jointly optimized with two constraints to preserve feature-level and structure-level consistency, respectively.

(b) We propose a dynamic attention weighted GCNs (DAW-GCNs) module to capture spatiotemporal dependencies in each level's graph representation by integrating GCNs with learnable weights on features from adjacent time windows. Furthermore, a temporal attention module learns weights for node features at each level in each time window, enabling the model to identify discriminative time windows and capture the dynamic nature of each level's brain network. Finally, a cross-layer attention module enriches LL representations with GL context via multi-head attention mechanisms, facilitating information exchange between feature representations at different levels.

(c) Our framework integrates phenotypic information into the output features to improve classification performance. It aims to identify multi-scale biomarkers, including ROI-level connections and ROI-to-global network connections. This enhances interpretability for brain disorder diagnosis.

2 Related Work

2.1 Hierarchical GNNs for Brain Network Representation Learning

The brain exhibits inherent hierarchical organization, spanning from individual ROIs to functional subnetworks and large-scale networks. This architecture supports complex cognitive processes through nested functional modules. Motivated by this, hierarchical GNN architectures are explored to model brain networks at multiple spatial resolutions (Li et al., 2025b). More recent work has introduced hierarchical graph transformers that simulate information flow from local regions to global communities. BrainHGT employs hierarchical attention mechanisms to capture cross-scale interactions, achieving competitive performance on brain disorder diagnosis (Ma et al., 2026b). Hierarchical pooling strategies have also been adapted to progressively coarsen graphs and extract compact representations (Xu et al., 2024; Jiang et al., 2024). These methods demonstrate the value of explicit hierarchical modeling for brain network analysis by reducing computational complexity and preserving task-relevant information. MHNet proposed a hierarchical framework that explicitly constructs three levels of brain networks for diagnosing neurodevelopmental disorders (Li et al., 2025d). However, the hierarchical levels in MHNet are processed independently through separate branches, and their outputs are only concatenated at the final fusion stage without explicit cross-level information exchange.

Existing hierarchical GNNs for brain disorder diagnosis suffer from common limitations. They often rely on fixed aggregation strategies, such as average pooling or predefined hierarchical partitions based on anatomical atlases (Fan et al., 2025). This limits their flexibility in capturing disorder-induced changes in functional network organization. On the other hand, most models process hierarchical levels independently and lack explicit cross-level interaction, preventing them from modeling the bidirectional modulation between local and global representations (Xue et al., 2025). To overcome these limitations, our model learns hierarchical aggregation weights directly from data and incorporates cross-level attention. This enables adaptation to disorder-specific brain network reorganization.

2.2 Spatio-Temporal Graph Learning for Dynamic Brain Networks

Recurrent and convolutional sequence models couple GNNs with RNNs, LSTM networks, or temporal convolutional networks (TCNs) to process sequences of dynamic connectivity matrices (Tang et al., 2025). These methods model spatial relationships within each time window and temporal transitions across windows, capturing transient disruptions that are invisible in averaged connectivity. For instance, TDNet employs a TCN with dilated convolutions to capture long-range temporal patterns from rs-fMRI for ADHD identification (Zhu et al., 2024).

Attention-based spatio-temporal models employ multi-head self-attention or Transformer architectures to jointly model spatial and temporal dependencies. STAGIN learns dynamic graph representations with both spatial and temporal attention, enabling simultaneous identification of discriminative time windows and brain regions for ASD and ADHD diagnosis (Kim et al., 2021). STHAG introduces a dynamic adaptive-neighbor GCN with temporal sliding self-attention for full-scale temporal dependency extraction (Liu et al., 2024b), while STIGR extends hybrid modeling to multiple neuroimaging modalities through local dynamic graph convolution, global spatio-temporal attention, and contrastive-learning-based adjacency matrix learning (Liu et al., 2025a). More recent works including BrainDGT (Shehzad et al., 2025a) and BrainMGT (Shehzad et al., 2025b), achieve strong performance by capturing long-range dependencies. However, they often integrate static and dynamic connectivity through late fusion or temporal averaging, which fails to fully exploit the complementary nature of the two modalities. Moreover, the interpretability of learned dynamic features remains an open challenge.

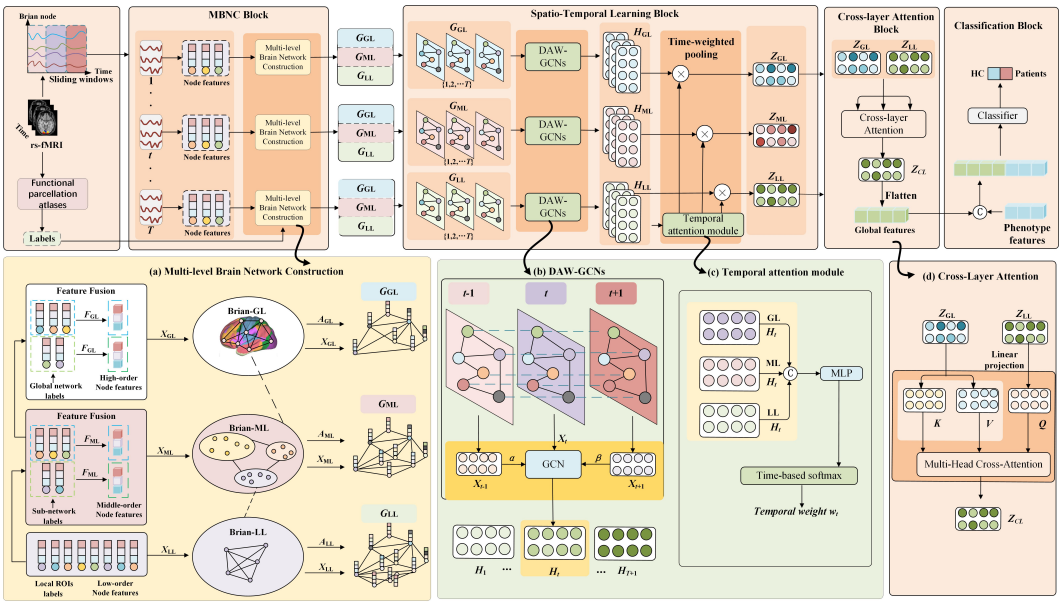


Figure 1: The overall framework of the proposed HADBN.

3 METHOD

The overall framework of the proposed HADBN is presented in Fig. 1. First, we construct hierarchical graph representations from rs-fMRI data at three progressive levels in multi-level brain network construction (MBNC) block. Second, a spatiotemporal learning block, including DAW-GCNs module and time-weighted pooling, is designed to learn spatiotemporal dependencies in each level's graph representation. Then, a cross-layer attention block facilitates information exchange between feature representations of different levels. Finally, phenotypic information and output features are fused for the classification of health controls (HCs) and patients.

3.1 MBNC BLOCK

To capture the multi-scale organization of brain networks, we construct a three-level hierarchical architecture guided by the prior Yeo atlas (Dickstein et al., 2006), including Brain-LL , Brain-ML, and Brain-GL. Specifically, the global large-scale networks at Brain-GL consists of seven brain subnetworks, i.e., visual network (VN), somatomotor network (SMN), dorsal attention network (DAN), ventral attention network (VAN), limbic network (LN), frontoparietal network (FPN), and default mode network (DMN). The middle subnetworks at Brain-ML encodes each of the seven brain subnetworks. For example, DMN include the frontal lobe (FL) and temporal lobe (TL). The local ROIs at Brain-LL encodes each of the middle subnetworks, such as TL is divided into regions Temporal-Sup-L, Temporal-Sup-R, Temporal-Mid-L, Temporal-Mid-R, and Temporal-Inf-L.

At Brain-LL, N_{L} ROIs for a subject are extracted using AAL atlas. The preprocessed BOLD time series from N_{L} ROIs are first divided into T overlapping windows using a sliding-window with length L and stride ST. Then, a time-series matrix $S^{(t)} \in \mathbb{R}^{N_{\mathrm{L}} \times L}$ is formed at a window $t \in \{1, 2, ..., T\}$ of a subject, which is used to compute a Pearson correlation matrix $R^{(t)} \in \mathbb{R}^{N_{\mathrm{L}} \times N_{\mathrm{L}}}$ between brain regions at the tth window. Thus, $R \in \mathbb{R}^{T \times N_{\mathrm{L}} \times N_{\mathrm{L}}}$ represents the dFC of a subject.

For the graph representation of LL, we define a brain graph $G_{\mathrm{LL}}^{(t)} = (A_{\mathrm{LL}}^{(t)}, X_{\mathrm{LL}}^{(t)})$ at the tth window, where the nodes denote ROIs and edges are the functional connections beween brain regions. The adjacency matrix $A_{\mathrm{LL}}^{(t)} \in [0, 1]^{N_{\mathrm{L}} \times N_{\mathrm{L}}}$ is the sparse dFC matrix, while node features $X_{\mathrm{LL}}^{(t)} \in \mathbb{R}^{N_{\mathrm{L}} \times N_{\mathrm{L}}}$ are the correlation vectors from $R^{(t)}$.

For the graph representation of ML, we define a brain graph $G_{\mathrm{ML}}^{(t)} = (A_{\mathrm{ML}}^{(t)}, X_{\mathrm{ML}}^{(t)})$ at the tth window, where the nodes denote the middle subnetworks and edges are the functional connections beween

those middle subnetworks. The adjacency matrix $A_{\text{ML}}^{(t)} \in [0,1]^{N_M \times N_M}$ is the RV coefficient matrix, where N_M denotes the number of subnetworks in ML. RV coefficient measures the strength of functional connections between two subnetworks by

$$RV(U,V) = \frac{Tr(UU^\top VV^\top)}{\sqrt{Tr[(UU^\top)^2]\,Tr[(VV^\top)^2]}}, \tag{1}$$

where $U \in \mathbb{R}^{L\times p}$ and $V \in \mathbb{R}^{L\times q}$ denote the BOLD time-series matrices of two sub-networks at the window t. p and q denote the number of ROIs contained in the two sub-networks, respectively. $Tr(\cdot)$ denotes matrix trace.

At ML, the nodes need to aggregate from LL. For that, we introduce learnable F-matrices to perform weighted aggregation under semantic constraints. $F_{\text{ML}}^{(t)} \in \mathbb{R}^{N_M \times m_i}$ is defined for the ith subnetwork, where m_i is the number of ROIs included in the ith subnetwork. Then, the node features $X_{\text{ML}}^{(t)} \in \mathbb{R}^{N_M \times N_{\text{L}}}$ of graph $G_{\text{ML}}^{(t)}$ are obtained by $X_{\text{ML}}^{(t)} = F_{\text{ML}}^{(t)} \tilde{X}_{\text{LL}}^{(t)}$, where $\tilde{X}_{\text{LL}}^{(t)} \in \mathbb{R}^{m_i \times N_{\text{L}}}$ is the feature sets of m_i nodes at LL.

Similarly, we define a brain graph $G_{\text{GL}}^{(t)} = (A_{\text{GL}}^{(t)}, X_{\text{GL}}^{(t)})$ at the tth window for graph representation of GL. The adjacency matrix $A_{\text{GL}}^{(t)} \in [0,1]^{N_G \times N_G}$ is the RV coefficient matrix, where N_G denotes the number of subnetworks in GL. The node features $X_{\text{GL}}^{(t)} \in \mathbb{R}^{N_G \times N_{\text{L}}}$ with $X_{\text{GL}}^{(t)} = F_{\text{GL}}^{(t)} \tilde{X}_{\text{ML}}^{(t)}$, where $F_{\text{GL}}^{(t)} \in \mathbb{R}^{N_G \times g_i}$ is defined for the ith subnetwork, where g_i is the number of subnetworks included in the ith network of GL. $\tilde{X}_{\text{ML}}^{(t)} \in \mathbb{R}^{g_i \times N_{\text{L}}}$ is the feature sets of g_i nodes at ML.

The learnable F-matrices ($F_{\text{ML}}^{(t)}$ and $F_{\text{GL}}^{(t)}$) are initialized as a uniform distribution and constrained to be non-negative with unit sum via ReLU activation and normalization. Moreover, to ensure aggregation fidelity, we impose two constraints on the F-matrices to preserve feature-level and structure-level consistency, respectively. For feature-level consistency, a cosine similarity loss enforces the aggregated features remaining close to the mean of their source features by

$$\mathcal{L}_{\text{feat},l}^{(t)} = \Big(1 - \frac{1}{N_l}\sum_{j=1}^{N_l} \cos\Big(X_{l,j}^{(t)},\ \mu_{\text{source}}^{(t)}\Big), \tag{2}$$

where $\mu_{\text{source}}^{(t)}$ is the mean of node features $\tilde{X}_{\text{source}}^{(t)}$ in source level. $l \in \{\text{ML}, \text{GL}\}$, $N_l \in \{N_M, N_G\}$, and source $\in \{\text{LL}, \text{ML}\}$. For structure-level consistency, structural loss $\mathcal{L}_{F,l}$ regularizes the F-vectors to be consistent with the hierarchical organization of the brain by $\mathcal{L}_F$ with is a weighted combination of five regularization terms: an L2 regularizer, an L1 sparsity term, a diversity term, an information-preservation term, and a row-sum constraint. The detail expressions of these terms are given in Appendix A. The feature-level loss $\mathcal{L}_{\text{feat}} = \frac{1}{T}\sum_{t=1}^{T}(\mathcal{L}_{\text{feat,ML}}^{(t)} + \mathcal{L}_{\text{feat,GL}}^{(t)})$ and structure-level loss $\mathcal{L}_F = \mathcal{L}_{F,\text{ML}} + \mathcal{L}_{F,\text{GL}}$ encourage the F-matrices to be sparse, diverse, and interpretable while preserving both local feature fidelity and global structural information. This design enhances the representation ability and the prior-guided interpretability of each level.

Finally, we introduce a region-level diversity loss $\mathcal{L}_{\text{div}}$ at the LL to promote functional specialization across large-scale networks of GL, This loss encourages the framework to learn dissimilar feature representations for local ROIs divided into distinct large-scale networks of GL by

$$\mathcal{L}_{\text{div}} = \frac{1}{M}\sum_{a\neq b}\ \sum_{i\in \text{GL}_a, j\in \text{GL}_b} \big|\cos(z_i, z_j)\big|, \tag{3}$$

where z_i is the time-averaged feature vector of $z_i^{(t)} = X_{\text{LL},i}^{(t)}$ for region i. GL_a and GL_b denote two different large-scale networks of GL, and M is the number of cross-GL ordered region pairs. Minimizing $\mathcal{L}_{\text{div}}$ separates functionally distinct networks in the feature space to enhance both interpretability and classification performance.

3.2 Spatial-Temporal Learning Block

3.2.1 Dynamic Attention Weighted GCNs

DAW-GCNs extends standard GCNs by incorporating temporal context from adjacent windows. Specifically, for $G_l^{(t)}$ $l \in \{\text{LL}, \text{ML}, \text{GL}\}$ level, the spatial graph convolution is defined as

$$H_l^{(t)} = \sigma(\hat{A}_l^{(t)} X_l^{(t)} W + \alpha_l^{(t)} \odot X_l^{(t-1)} W + \beta_l^{(t)} \odot X_l^{(t+1)} W), \tag{4}$$

where $\hat{A}_l^{(t)} = D^{-1/2}(A_l^{(t)} + I)D^{-1/2}$is the symmetrically normalized adjacency matrix with self-loops; $\alpha_l^{(t)} = g_\alpha(X_l^{(t)}, X_l^{(t-1)}, X_l^{(t+1)}; \Theta_\alpha)$ and $\beta_l^{(t)} = g_\beta(X_l^{(t)}, X_l^{(t-1)}, X_l^{(t+1)}; \Theta_\beta)$ are the learnable time attention coefficients dynamically generated from the node features, where g_α, g_β are functions with learnable parameters $\Theta_\alpha, \Theta_\beta$, respectively; $\odot$ performs broadcast element-wise multiplication. For boundary windows, the temporal terms are masked out. After two layer DAW-GCNs, $H_l^{(t)} \in \mathbb{R}^{N_l \times d_l}$ with feature dimension d_l at l level is input to the time-weighted pooling for identifying the informative temporal windows.

3.2.2 Temporal Attention Module

We compute temporal importance weights from multi-level representations rather than from raw signals or single-level features. Specifically, the flattened $H_l^{(t)}$ of three levels are concatenated along the feature dimension. Then, they pass through a multilayer perceptron (MLP) with Tanh activations. The resulting logits are then normalized via softmax with a learnable temperature parameter τ, yielding a temporal weight vector $W = \{w_t | t = 1, 2, ..., T\}$. Further, to focus on discriminative time windows rather than distributing attention uniformly, we impose a temporal entropy loss by

$$\mathcal{L}_{\text{temp}} = -\sum_{t=1}^{T} w_t \log(w_t + \epsilon), \tag{5}$$

where ϵ is a small constant for numerical stability. Finally, time-weighted pooling aggregates features across the temporal dimension using the learned weights, yielding node-level representations $Z_l = W H_l$ ($H_l \in \mathbb{R}^{T \times N_l \times d_l}$). Thus, the temporal attention is directly conditioned on comprehensive multi-scale spatial features for the specific disorder.

3.3 Cross-Layer Attention Block

A cross-layer attention module is used to enrich LL representations with GL context. Specifically, the features $Z_{\text{LL}} \in \mathbb{R}^{N_{\text{L}} \times d_{\text{L}}}$ are first projected to $\tilde{Z}_{\text{LL}} \in \mathbb{R}^{N_{\text{L}} \times d_{\text{G}}}$. The query, key, and value matrices are computed as

$$Q = \tilde{Z}_{\text{LL}} W_Q, \quad K = Z_{\text{GL}} W_K, \quad V = Z_{\text{GL}} W_V, \tag{6}$$

where $W_Q, W_K, W_V \in \mathbb{R}^{d_{\text{G}} \times d_{\text{G}}}$ are learnable projection matrices. For each attention head $h \in \{1, 2, ..., H\}$, the scaled dot-product attention is computed as

$$\text{head}_h = \text{softmax}\left(\frac{Q_h K_h^\top}{\sqrt{d_k}}\right) V_h, \tag{7}$$

where $d_k = d_{\text{G}}/H$ is the dimension of each head. Q_h, K_h, V_h denote the hth feature slices of Q, K, V along the feature dimension, respectively. The outputs of all heads are concatenated and projected by an output matrix $W_O \in \mathbb{R}^{d_{\text{G}} \times d_{\text{G}}}$. This process allows each brain region to attend to the most relevant large-scale network contexts, effectively integrating fine-grained regional information with high-level network representations. Finally, $Z_{\text{CL}} \in \mathbb{R}^{N_{\text{LL}} \times d_{\text{G}}}$ is flattened and output to classification block as the discriminative representation.

3.4 Phenotype-Enhanced Classification

Phenotypic information characterizes the global physiological, genetic, or clinical profiles of subjects, exhibiting inherent orthogonality and complementarity with the spatiotemporal brain features

Table 1: Summary of ADHD-200 and ABIDE datasets.

Dataset	Gender (F/M)	Age (mean ± std)	Class	Subjects
ABIDE-I	152/873	16.5 ± 7.4	HCs	537
			ASD	488
ADHD-200	180/361	12.0 ± 2.9	HCs	298
			ADHD	243

extracted by Z_{CL}. The integration of the two modalities facilitates the enhancement of discriminative representations, thereby improving classification performance and interpretability. The phenotypic information (e.g., age, IQ, sex, handedness) and the flattened node embeddings Z_{CL} are independently projected, concatenated, and subsequently fed into a classifier for classification. The classification loss is the binary cross-entropy defined as $\mathcal{L}_{\text{CE}}$.

In summary, the total loss function of the framework can be formulated as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda_1 \mathcal{L}_{\text{div}} + \lambda_2 \mathcal{L}_{\text{feat}} + \lambda_3 \mathcal{L}_F + \lambda_4 \mathcal{L}_{\text{temp}}, \tag{8}$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are hyperparameters balancing the contributions of each loss.

4 Experiments

4.1 Datasets and Data Preprocessing

ADHD-200: The ADHD-200 datasets provide multi-site rs-fMRI data for ADHD. Considering quality-controlled imaging data, we choose the data from four sites, namely New York University Medical Center (NYU), Peking University (PU), Kennedy Krieger Institute (KKI), and NeuroImage (NI). The subject information is shown in Table 1. BOLD signals of these subjects are obtained from the Connectome Project website (http://www.preprocessed-connectomes-project.org/adhd200).

ABIDE-I: The ABIDE-I datasets are from 17 international sites for autism spectrum disorder (ASD). The subject information is shown in Table 1. BOLD signals of these subjects are obtained from the the Preprocessed Connectomes Project (PCP), which is available at http://preprocessed-connectomes-project.org/abide/.

After preprocessing, the BOLD signals were parcellated into 116 cerebral regions using the AAL atlas. We selected 90 cerebral cortical and subcortical regions to focus on the primary functional networks implicated in ASD and ADHD.

4.2 Experimental Settings

To ensure consistency across subjects with different scan durations, we fixed T by sampling a fixed number of windows. To avoid abrupt transitions between adjacent windows, we selected a window length of 40 TR and a stride of 10 TR, which achieved the highest adjacent-window similarity (0.7595) and a moderate top 20% connection similarity (0.5059), providing a good balance between temporal consistency and dynamic sensitivity. The detail of selecting window parameters is given in Appendix B.

We set $N_{\text{L}} = 90$, $N_{\text{M}} = 17$ and $N_{\text{G}} = 7$ for three levels based on AAL atlas and Yeo atlas. The dimensions of graph representation features are $d_{\text{L}} = 32$, $d_{\text{M}} = 16$ and $d_{\text{G}} = 16$ for the three levels. The model is trained for 100 epochs with a batch size of 128, using the Adam optimizer. A warmup-cosine learning rate schedule is adopted. The loss weights are set as $\lambda_1 = 0.1$, $\lambda_2 = 0.05$, $\lambda_3 = 1.0$, and $\lambda_4 = 0.01$.

We adopt 10-fold cross-validation to evaluate the classification performance and generalizability of the proposed framework. The cross-validation procedure is repeated five times and the averaged results are reported. Classification performance is measured by accuracy (ACC), area under the receiver operating characteristic curve (AUC), sensitivity (SEN), and specificity (SPE).

Table 2: Classification performance comparison on the ABIDE-I and ADHD-200 datasets.

	ABIDE-I		ADHD-200	
Model	ACC (%)	AUC (%)	ACC (%)	AUC (%)
RF	55.25±2.71	55.39	57.15±2.39	56.33
SVM	63.74±3.99	**69.90**	57.96±2.60	59.40
LSTM	64.77±4.12	68.84	62.10±3.84	64.24
BrainGNN (2021)	**69.31±3.03**	69.52	66.63±5.41	64.03
STGCN (2020)	64.20±3.70	62.82	67.61±4.20	68.59
STAGIN (2021)	65.52±3.12	65.62	68.23±3.29	68.21
STHAG (2024)	66.23±3.40	68.82	69.14±1.83	69.50
STIGR (2025)	65.63±3.12	68.50	69.03±1.84	**78.11**
HADBN (Ours)	68.14±0.54	69.32	**70.78±0.34**	73.48

4.3 Classification Performance

To evaluate the effectiveness of the proposed HADBN, we compare it against four categories of baselines: conventional machine learning methods (RF and SVM), a sequential deep model (LSTM), a brain-specific GNN (BrainGNN), and state-of-the-art spatio-temporal GNNs (STGCN, STAGIN, ST-HAG, and STIGR). All baseline models were implemented using their publicly available source code and evaluated under the optimal hyperparameter settings reported in their original publications to ensure a fair comparison.

The classification results on the ABIDE-I and ADHD-200 datasets are presented in Table 2. Among the spatio-temporal GNNs, STGCN, STAGIN, STHAG, and STIGR show progressively better performance than RF, SVM and LSTM, highlighting the advantage of jointly modeling spatial and temporal dynamics. In particular, STHAG and STIGR represent the current frontier in spatio-temporal brain network modeling and achieve the highest accuracy on both datasets. As a result, our proposed HADBN achieves the highest performance on ADHD-200 among all compared methods, reaching 68.14% and 70.78% accuracy on ABIDE-I and ADHD-200, respectively, with competitive AUC values. These results demonstrate that explicitly modeling hierarchical spatial organization and principled spatio-temporal integration leads to more robust and accurate brain disorder diagnosis.

4.4 Ablation Study

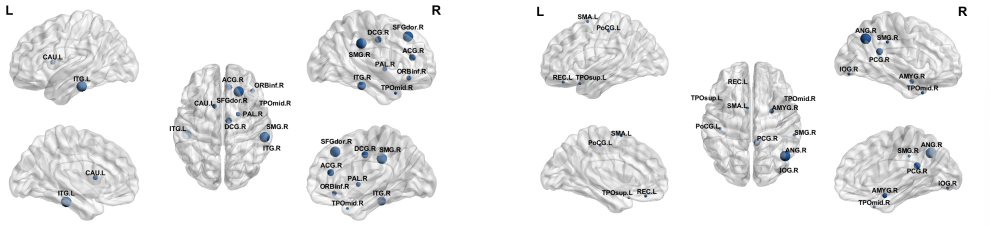
To evaluate the contribution of each core component in HADBN, we conduct ablation experiments on both datasets in Table 3. On ADHD-200, removing cross-attention causes the most severe degradation, with accuracy falling to 66.15% and sensitivity dropping to 57.58%. This highlights the essential role of bidirectional information exchange between fine-grained LL features and global GL context for capturing multi-scale ADHD-related patterns. The w/o temporal attention variant shows a moderate decline (accuracy 66.34%), suggesting that while adaptive temporal weighting contributes to identifying discriminative dynamic states, the DAW-GCNs' cross-window fusion partially compensates for its removal. The w/o F-weights variant achieves the highest accuracy among ablated models (69.48%), indicating that hierarchical pooling, though beneficial, has a less critical impact than cross-level integration. On ABIDE-I, a similar pattern is observed. These findings validate the effectiveness of our hierarchical and dynamic modeling framework, where each module addresses a distinct aspect of the brain's multi-scale spatio-temporal organization.

4.5 Biomarker Analysis

We visualize the learned F-matrix weights for each level pooling to reveal which brain regions contribute most to the final classification. To complement the network-level F-matrix findings, we compute node-level importance scores from the features of LL, identifying the top 10 brain regions contributing to brain disorder diagnosis. The 10-top discriminative regions are listed for ADHD and ASD in Appendix C.

Table 3: Ablation study results on ABIDE-I and ADHD-200 datasets (10-fold CV, mean %).

Dataset	Model	ACC (%)	SEN (%)	SPE (%)	AUC (%)
ABIDE-I	w/o F-weights	65.08	59.45	70.27	65.66
	w/o cross-attention	65.75	60.44	70.61	67.13
	w/o temporal attention	63.32	60.68	65.70	63.60
	w/o α and β	64.49	66.86	62.40	64.57
	Full model	68.14	69.21	64.54	69.32
ADHD-200	w/o F-weights	69.48	67.17	71.52	69.56
	w/o cross-attention	66.15	57.58	73.21	66.55
	w/o temporal attention	66.34	70.05	63.44	66.38
	w/o α and β	67.27	65.97	68.52	67.42
	Full model	70.78	69.15	71.42	73.48



(a) ADHD **(b)** ASD

Figure 2: Top 10 discriminative brain regions for ADHD and ASD.

For ADHD classification, the F-vectors show that the 10-top discriminative regions shown in Fig. 2 are predominantly distributed across the DMN, SN, and LN. The identified discriminative regions align with established neuroimaging literature (Cortese et al., 2012). The prominent involvement of the right superior frontal gyrus and anterior cingulate cortex is consistent with meta-analytic evidence of fronto-striatal and fronto-parietal circuit dysfunctions in ADHD, supporting the role of these regions in sustained attention and response inhibition deficits (Dickstein et al., 2006). The right supramarginal gyrus and caudate nucleus have been repeatedly implicated in attentional reorienting and cognitive flexibility impairments in ADHD populations (Hart et al., 2013). The basal ganglia structures (globus pallidus and caudate) align with well-documented abnormalities in striatal reward processing and motor control circuits, which are central to the pathophysiology of ADHD (Hoogman et al., 2017).

For ASD classification, the 10-top discriminative regions are shown in Fig. 2. The discriminative regions in the prominence of DMN and limbic regions align with the socio-emotional and communicative deficits that characterize ASD (Lynch et al., 2013). Consistent with prior findings, the strong contribution of the angular gyrus, posterior cingulate cortex, and amygdala is consistent with the social cognition and emotion processing deficits that are core features of ASD (Padmanabhan et al., 2017; Baron-Cohen et al., 2000). The involvement of the right inferior occipital gyrus and superior temporal gyrus reflects the visual processing and language comprehension difficulties frequently observed in ASD (Donovan & Basson, 2017).

5 Conclusion

We presented HADBN framework for brain disorder diagnosis, which combines a three-level spatial architecture. The framework enables data-driven hierarchical pooling and bidirectional information flow between local regions and global networks. HADBN achieved better accuracy for ASD and ADHD diagnosis, respectively. It alse identified disorder-relevant brain regions that align with prior neuroimaging findings, offering interpretable biomarkers for ADHD and ASD.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing of selected paragraphs and for code-level boilerplate generation during prototyping. We also used AI to retrieve some of the relevant literature, and every reference retrieved by AI was manually confirmed for authenticity and accuracy. We have not used generative AI tools for the core research design, the derivation of mathematical formulations, the experimental analysis, or the interpretation of results. We have reviewed all AI-assisted work: every generated sentence was manually edited and verified for technical accuracy, and all AI-suggested code was tested and rewritten as needed. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This study uses two publicly available, de-identified neuroimaging datasets (ABIDE-I and ADHD-200). No new data were collected from human participants. All preprocessing and analyses were performed on summary statistics (functional connectivity matrices) and aggregate measures, in accordance with the data use agreements of the original studies. No personally identifiable information is reported. The work does not raise additional ethics concerns beyond those addressed in the source datasets.

REPRODUCIBILITY STATEMENT

To support reproducibility, we provide a detailed description of the model architecture (Section 3), training procedure (Section 4.2), and evaluation protocol (10-fold cross-validation repeated five times with different random seeds). Hyperparameter settings are listed explicitly, and dataset preprocessing follows standardized pipelines. Upon acceptance, we will release the source code and a reproducibility script under an open-source license as supplementary material. Pretrained model weights and processed connectivity matrices will also be released for non-reproducing verification.

REFERENCES

Simon Baron-Cohen, Howard A. Ring, Edward T. Bullmore, Sally Wheelwright, Chris Ashwin, and Steve C. R. Williams. The amygdala theory of autism. *Neuroscience & Biobehavioral Reviews*, 24(3):355–364, 2000. ISSN 0149-7634. doi: 10.1016/S0149-7634(00)00011-7.

Bharat B Biswal and Lucina Q Uddin. The history and future of resting-state functional magnetic resonance imaging. *Nature*, 641(8065):1121–1131, 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-08953-9.

Samuele Cortese, Clare Kelly, Camille Chabernaud, Erika Proal, Adriana Di Martino, Michael P. Milham, and F. Xavier Castellanos. Toward systems neuroscience of adhd: a meta-analysis of 55 fmri studies. *American Journal of Psychiatry*, 169(10):1038–1055, 2012. ISSN 0002-953X. doi: 10.1176/appi.ajp.2012.11101521.

Steven G. Dickstein, Katie Bannon, F. Xavier Castellanos, and Michael P. Milham. The neural correlates of attention deficit hyperactivity disorder: an ale meta-analysis. *Journal of Child Psychology and Psychiatry*, 47(10):1051–1062, 2006. ISSN 0021-9630. doi: 10.1111/j.1469-7610.2006.01671.x.

Alex P. A. Donovan and M. Albert Basson. The neuroanatomy of autism - a developmental perspective. *Journal of Anatomy*, 230(1):4–15, 2017. ISSN 0021-8782. doi: 10.1111/joa.12542.

Y Du, S Fang, X He, and V D Calhoun. A survey of brain functional network extraction methods using fmri data. *Trends in Neurosciences*, 47(8):608–621, 2024. ISSN 0166-2236. doi: 10.1016/j.tins.2024.05.011.

Zhengyuan Fan, Hongfei Ji, Jie Li, Jie Zhuang, and Qian Qian. Hibignn: Hierarchical bilateral graph neural network for fmri analysis. *Computing and Informatics*, 44(5):1040–1070, 2025. ISSN 1335-9150. doi: 10.31577/cai_2025_5_1040.

Heledd Hart, Joaquim Radua, Tomohiro Nakao, David Mataix-Cols, and Katya Rubia. Meta-analysis of functional magnetic resonance imaging studies of inhibition and attention in attention-deficit/hyperactivity disorder: exploring task-specific, stimulant medication, and age effects. *JAMA Psychiatry*, 70(2):185–198, 2013. ISSN 2168-622X. doi: 10.1001/jamapsychiatry.2013.277.

Martine Hoogman, Janita Bralten, Derrek P. Hibar, Maarten Mennes, Marcel P. Zwiers, Lizanne S. J. Schweren, Kimm J. E. van Hulzen, Sarah E. Medland, Elena Shumskaya, Neda Jahanshad, et al. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *The Lancet Psychiatry*, 4(4):310–319, 2017. ISSN 2215-0366. doi: 10.1016/S2215-0366(17)30049-4.

Yu Jiang, Zhibin He, Zhihao Peng, and Yixuan Yuan. Hierarchical graph learning with small-world brain connectomes for cognitive prediction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, pp. 401–410, 2024. doi: 10.1007/978-3-031-72086-4_38.

Byung-Hoon Kim, Jong Chul Ye, and Jae-Jin Kim. Learning dynamic graph representation of brain connectome with spatio-temporal attention. In *Advances in Neural Information Processing Systems*, volume 34, pp. 4314–4327, 2021.

Timothy O Laumann, Abraham Z Snyder, and Caterina Gratton. Challenges in the measurement and interpretation of dynamic functional connectivity. *Imaging Neuroscience*, 2:1–21, 2024. doi: 10.1162/imag_a_00366.

Chaojun Li, Kai Ma, Shengrong Li, Xiangshui Meng, Ran Wang, Daoqiang Zhang, and Qi Zhu. Multi-channel spatio-temporal graph attention contrastive network for brain disease diagnosis. *NeuroImage*, 307:121013, 2025a. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2025.121013.

Qiuyan Li, Guangming Li, Mingxin Liu, et al. Multi-scale atlases based hierarchical heterogeneous graph learning for brain disorder diagnosis. *Expert Systems with Applications*, 305:130752, 2025b. ISSN 0957-4174. doi: 10.1016/j.eswa.2025.130752.

Yueyang Li, Weiming Zeng, Wenhao Dong, Luhui Cai, Lei Wang, Hongyu Chen, Hongjie Yan, Lingbin Bian, and Nizhuan Wang. Mhnet: Multi-view high-order network for diagnosing neurodevelopmental disorders using resting-state fmri. *Journal of Imaging Informatics in Medicine*, 2025c. ISSN 2948-2933. doi: 10.1007/s10278-025-01413-w.

Yueyang Li, Weiming Zeng, Wenhao Dong, Luhui Cai, Lei Wang, Hongyu Chen, Hongjie Yan, Lingbin Bian, and Nizhuan Wang. Mhnet: Multi-view high-order network for diagnosing neurodevelopmental disorders using resting-state fmri. *Journal of Imaging Informatics in Medicine*, 2025d. ISSN 2948-2933. doi: 10.1007/s10278-025-01413-w.

Mianxin Liu, Han Zhang, Feng Shi, and Dinggang Shen. Hierarchical graph convolutional network built by multiscale atlases for brain disorder diagnosis using functional connectivity. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11):15182–15194, 2024a. ISSN 2162-237X. doi: 10.1109/TNNLS.2023.3282961.

Rui Liu, Zhi-An Huang, Yao Hu, Lei Huang, Ka-Chun Wong, and Kay Chen Tan. Spatio-temporal hybrid attentive graph network for diagnosis of mental disorders on fmri time-series data. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(6):4046–4058, 2024b. ISSN 2471-285X. doi: 10.1109/TETCI.2024.3386612.

Rui Liu, Zhi-An Huang, Yao Hu, Lei Huang, Ka-Chun Wong, and Kay Chen Tan. Spatio-temporal hybrid attentive graph network for diagnosis of mental disorders on fmri time-series data. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(6):4046–4058, 2024c. ISSN 2471-285X. doi: 10.1109/TETCI.2024.3386612.

Rui Liu, Yao Hu, Jibin Wu, Ka-Chun Wong, Zhi-An Huang, Yu-An Huang, and Kay Chen Tan. Dynamic graph representation learning for spatio-temporal neuroimaging analysis. *IEEE Transactions on Cybernetics*, 55(3):1121–1134, 2025a. ISSN 2168-2267. doi: 10.1109/TCYB.2025.3531657.

Rui Liu, Yao Hu, Jibin Wu, Ka-Chun Wong, Zhi-An Huang, Yu-An Huang, and Kay Chen Tan. Dynamic graph representation learning for spatio-temporal neuroimaging analysis. *IEEE Transactions on Cybernetics*, 55(3):1121–1134, 2025b. ISSN 2168-2267. doi: 10.1109/TCYB.2025.3531657.

Charles J. Lynch, Lucina Q. Uddin, Kaustubh Supekar, Amirah Khouzam, Jennifer Phillips, and Vinod Menon. Default mode network in childhood autism: posteromedial cortex heterogeneity and relationship with social deficits. *Biological Psychiatry*, 74(3):212–219, 2013. ISSN 0006-3223. doi: 10.1016/j.biopsych.2012.12.013.

Jiajun Ma, Yongchao Zhang, Chao Zhang, Zhao Lv, and Shengbing Pei. Brainhgt: A hierarchical graph transformer for interpretable brain network analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 17617–17625, 2026a. doi: 10.1609/aaai.v40i21.38817.

Jiajun Ma, Yongchao Zhang, Chao Zhang, Zhao Lv, and Shengbing Pei. Brainhgt: A hierarchical graph transformer for interpretable brain network analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 17617–17625, 2026b. doi: 10.1609/aaai.v40i21.38817.

Aarthi Padmanabhan, Charles J. Lynch, Marie Schaer, and Vinod Menon. The default mode network in autism. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 2(6):476–486, 2017. ISSN 2451-9022. doi: 10.1016/j.bpsc.2017.04.004.

Ahsan Shehzad, Dongyu Zhang, Shuo Yu, Shagufta Abid, and Feng Xia. Dynamic graph transformer for brain disorder diagnosis. *IEEE Journal of Biomedical and Health Informatics*, 29(6):4388–4400, 2025a. ISSN 2168-2194. doi: 10.1109/JBHI.2025.3538967.

Ahsan Shehzad et al. Multiscale graph transformer for brain disorder diagnosis. *IEEE Transactions on Consumer Electronics*, 71(2):5535–5546, 2025b. ISSN 0098-3063. doi: 10.1109/TCE.2025.3543297.

Jingye Tang, Tianqing Zhu, Wanlei Zhou, and Wei Zhao. Graph neural networks for fmri functional brain networks: A survey. *Neural Networks*, 194:108137, 2026. ISSN 0893-6080. doi: 10.1016/j.neunet.2025.108137.

Jun Tang, Jie Chen, Miaojun Hu, Yao Hu, Zixi Zhang, and Liuming Xiao. Diagnosis of autism spectrum disorder (asd) by dynamic functional connectivity using gnn-lstm. *Sensors*, 25(1):156, 2025. ISSN 1424-8220. doi: 10.3390/s25010156.

Xin Wen, Shijie Guo, Li Dong, Xiaobo Liu, Wenbo Ning, Jie Shi, Songhua Liu, Cheng Luo, Rui Cao, and Dezhong Yao. A consistency-driven pseudo-labeling framework for robust functional connectivity modeling in neuropsychiatric disorder diagnosis. *Engineering Applications of Artificial Intelligence*, 166(Part A):113556, 2026. ISSN 0952-1976. doi: 10.1016/j.engappai.2025.113556.

Ruiwei Xie, An Zeng, Dan Pan, and Yuqing Zhao. Cgfb-gnn: Explainable spatio-temporal gnn for cognition-guided functional brain network. *Biomedical Signal Processing and Control*, 126:108840, 2026. ISSN 1746-8094. doi: 10.1016/j.bspc.2025.108840.

Jiaxing Xu, Qingtian Bian, Xinhang Li, Aihu Zhang, Yiping Ke, Miao Qiao, Wei Zhang, Wei Khang Jeremy Sim, and Balázs Gulyás. Contrastive graph pooling for explainable classification of brain networks. *IEEE Transactions on Medical Imaging*, 43(9):3292–3305, 2024. ISSN 0278-0062. doi: 10.1109/TMI.2024.3394538.

Rundong Xue, Hao Hu, Zeyu Zhang, Xiangmin Han, Juan Wang, and Yue Gao. Dhgformer: Dynamic hierarchical graph transformer for disorder brain disease diagnosis. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, 2025.

Mengni Zhou, Miaofeng Wang, Rongkun Mi, Yan Niu, Xiaohong Cui, Ran Gao, Yuxi Li, Xin Wen, and Jie Xiang. Ms-stfnn: A multi-scale spatio-temporal fusion neural network for fmri-based depression diagnosis. *Neural Networks*, 199:108685, 2026. ISSN 0893-6080. doi: 10.1016/j.neunet.2026.108685.

Lingyao Zhu, Mingliang Wang, Xizhi Li, et al. Dynamic functional connectivity analysis with temporal convolutional network for attention deficit/hyperactivity disorder identification. *Frontiers in Neuroscience*, 17:1322967, 2024. ISSN 1662-453X. doi: 10.3389/fnins.2023.1322967.

APPENDIX A

$\mathcal{L}_{F,l}$ is a weighted combination of five constraint terms by

$$\mathcal{L}_{F,l} = \lambda_{\text{reg}}\mathcal{L}_{\text{reg}} + \lambda_{\text{sparse}}\mathcal{L}_{\text{sparse}} + \lambda_{F\text{div}}\mathcal{L}_{F\text{div}} + \lambda_{\text{preserve}}\mathcal{L}_{\text{preserve}} + \lambda_{\text{rowsum}}\mathcal{L}_{\text{rowsum}}, \tag{9}$$

where the weights are set to $\lambda_{\text{reg}} = 0.001$, $\lambda_{\text{sparse}} = 0.01$, $\lambda_{F\text{div}} = 0.05$, $\lambda_{\text{preserve}} = 0.02$, and $\lambda_{\text{rowsum}} = 0.01$. The five terms are defined as follows.

Let $\mathcal{F} = \{F_l^{(t)}[k,:] \mid t = 1,\ldots,T,\ k = 1,\ldots,N_l\}$ denotes the set of all F-vectors, where F-vector is $f = F_l^{(t)}[k,:]$. Thus, $|\mathcal{F}| = T \cdot N_l$. The L2 regularization term is

$$\mathcal{L}_{\text{reg}} = \frac{1}{|\mathcal{F}|}\sum_{f\in\mathcal{F}} \|f\|_2, \tag{10}$$

and the L1 sparsity term is

$$\mathcal{L}_{\text{sparse}} = \frac{1}{|\mathcal{F}|}\sum_{f\in\mathcal{F}} \|f\|_1. \tag{11}$$

Furthermore, the row-sum constraint encourages each F-vector to behave as a weighted average by

$$\mathcal{L}_{\text{rowsum}} = \frac{1}{|\mathcal{F}|}\sum_{f\in\mathcal{F}} \Big(\sum_j f_j - 1\Big)^2. \tag{12}$$

To encourage complementary features among aggregation units, the diversity term is computed as the average absolute cosine similarity between adjacent aggregated node-level representations by

$$\mathcal{L}_{F\text{div}} = \frac{1}{N_l - 1}\sum_{k=1}^{N_l-1} \big|\cos(z_k, z_{k+1})\big|, \tag{13}$$

where z_k is the time-averaged feature vector of $z_k^{(t)} = X_l^{(t)}[k,:] \in \mathbb{R}^{1\times N_l}$. The information preservation term enforces the aggregated features to retain the global characteristics of the source features by

$$\mathcal{L}_{\text{preserve}} = \max\big(0,\ 1 - \text{corr}(z_{\text{before}}, z_{\text{after}})\big), \tag{14}$$

where z_{before} and z_{after} are the graph-level vectors obtained by averaging the source features $\tilde{X}_{\text{source}}^{(t)}$ and aggregated features $X_{\text{source}}^{(t)}$, respectively, over the time and node dimensions. $\text{corr}(\cdot,\cdot)$ denotes the Pearson correlation coefficient.

APPENDIX B

We compared different window configurations by computing (i) the mean similarity between adjacent windows and (ii) the similarity of the top 20% strongest connections across windows. As summarized in Table 4, higher adjacent-window similarity indicates greater temporal consistency, whereas lower top 20% connection similarity suggests better retention of dynamic information. Therefore, we selected a window length of 40 TR and a stride of 10 TR, which achieved the highest adjacent-window similarity (0.7595) and a moderate top 20% connection similarity (0.5059), providing a good balance between temporal consistency and dynamic sensitivity.

APPENDIX C

The 10-top discriminative regions are listed in Table 5 for ADHD and ASD.

Table 4: Comparison of different sliding-window configurations.

ID	Window length (TR)	Stride (TR)	Windows	Adjacent similarity	Top-20% similarity
1	30	10	13	0.7559	0.4052
2	30	15	9	0.7398	0.3140
3	30	20	7	0.7317	0.2525
4	40	10	12	0.7595	0.5059
5	40	20	6	0.7200	0.3474
6	40	30	4	0.7051	0.2559
7	50	15	7	0.7392	0.4869
8	50	25	5	0.7074	0.3693

Table 5: Top 10 discriminative brain regions for ADHD and ASD.

ADHD

Rank	Brain region	Rank	Brain region
1	SFGdor.R	6	ACG.R
2	SMG.R	7	ORBinf.R
3	ITG.L	8	PAL.R
4	ITG.R	9	CAU.L
5	DCG.R	10	TPOmid.R

ASD

Rank	Brain region	Rank	Brain region
1	ANG.R	6	TPOmid.R
2	PCG.R	7	SMA.L
3	AMYG.R	8	IOG.R
4	PoCG.L	9	REC.L
5	SMG.R	10	TPOsup.L