SPIKING-TEMPORAL DIRECTED HIERARCHICAL PROPAGATION NETWORK FOR BRAIN DISORDER DIAGNOSIS

Anonymous authors
Paper under double-blind review

ABSTRACT

The brain organizes its functions hierarchically. Clarifying how local activity in brain disorders is transformed and propagated across functional hierarchies is essential for pursuing more precise neurobiological markers. Existing diagnosis methods typically construct graphs from inter-regional correlations or aggregate regions into higher-level representations, but they often decouple connectivity from ongoing regional dynamics and rely on undirected interactions, leaving the pathway from local temporal activity to system-level information flow insufficiently modeled. Thus, we propose Spiking-Temporal Directed Hierarchical Propagation Network (ST-DHPNet), a biologically inspired framework for multi-level brain network modeling for brain disorder diagnosis. ST-DHPNet first introduces a Spiking Temporal Brain Encoder that decomposes regional BOLD signals and processes consecutive observations through a stateful spiking encoder, producing activity-sensitive regional representations. It then employs an activity-guided spatial encoder that conditions functional-connectivity attention on regional activity states. Finally, a directional hierarchical propagation module aggregates regions into functional subnetworks using anatomical priors and performs effective-connectivity-guided propagation within and between subnetworks. This design explicitly couples regional temporal dynamics, activity-dependent spatial interactions, and directed subnetwork propagation. Experiments on three brain disorder datasets demonstrate performance improvements over other state-of-the-art methods in brain disorder diagnosis. ST-DHPNet reveals candidate biomarkers at three complementary levels: regional latent spike representations, inter-regional connectivity, and subnetwork coordination, providing an interpretable computational framework for brain disorder classification. The code is available at `https://anonymous.4open.science/r/ST-DHPNet`.

1 INTRODUCTION

The human brain exhibits a hierarchical organization across multiple functional scales, extending from neural activity within regions through inter-regional connectivity to functional subsystems (Sporns & Betzel, 2016). Such a multi-scale hierarchical structure plays a fundamental role in preserving the brain's functional homeostasis (Pines et al., 2022). Consequently, the pathophysiological mechanisms of neurodevelopmental and neurodegenerative disorders typically do not stem from abnormalities confined to a single hierarchical level, but instead involve systemic dysregulation across multiple scales (Rossignol & Frye, 2012; Cui et al., 2024). Such pathology may involve abnormal regional activity (Sudimac et al., 2022), weakened or aberrant connectivity (Corriveau-Lecavalier et al., 2024), and disrupted organization of functional systems (Li et al., 2022), all of which collectively impair the brain's cognitive functions.

Recent efforts to identify neurobiological biomarkers for brain disorders have primarily focused on functional connectivity (FC) analysis based on blood-oxygen-level-dependent (BOLD) functional magnetic resonance imaging (fMRI). This approach characterizes the coordinated activity patterns among brain regions by computing statistical associations between BOLD time series of predefined regions of interest (ROIs) (Kan et al., 2022). A large body of empirical evidence has demonstrated that FC abnormalities are closely associated with various neuropsychiatric disorders (Zheng et al., 2024). Concurrently, to identify localized neural dysfunction in critical areas, some studies adopt an ROI-wise strategy that directly analyzes regional BOLD time series (Zhang et al., 2022). These methods

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.

focus on the dynamic properties of individual ROIs. However, these two strategies are typically pursued in isolation. FC analysis emphasizes inter-regional interactions but often treats regional activity as static or stationary, thereby overlooking its intrinsic temporal dynamics. Conversely, ROI-wise analysis can effectively capture local neural abnormalities but lacks systematic modeling of how these regions are embedded within and influence the whole-brain hierarchical architecture.

Moreover, recent dynamic functional connectivity approaches have attempted to capture time-varying interactions among brain regions (Liu et al., 2023). However, most existing frameworks focus predominantly on functional connectivity and lack a complete hierarchical framework that extends from brain regional activity to the subnetwork level (Palop & Mucke, 2016).

To bridge this gap, we propose the Spiking-Temporal Directed Hierarchical Propagation Network (ST-DHPNet), which encodes regional brain activity through spiking temporal dynamics. It then employs a brain spatial encoder to characterize inter-regional dependencies and constructs a spatial hierarchy from regions to functional subnetworks. By capturing multi-scale abnormalities across regional latent activity, ROI-level connectivity, and subnetwork-level coordination, our framework provides a comprehensive and interpretable representation of brain dysfunction, offering a biologically grounded tool for understanding neurological and psychiatric disorders. The main contributions are as follows:

- We propose the Spiking-Temporal Directed Hierarchical Propagation Network (ST-DHPNet) for brain disorder diagnosis. ST-DHPNet provides a unified framework that links regional activity, activity-dependent spatial interactions, and subnetwork functional organization.
- A spiking temporal brain encoder and an activity-guided functional connectivity encoder are designed to model regional dynamics and activity-dependent spatial interactions.
- We design a directed hierarchical propagation module that explicitly models cross-scale information flow by leveraging effective connectivity and anatomical priors, thereby capturing how information is routed from individual regions to functional subnetworks.
- The experiments on three brain disorder datasets show that the proposed ST-DHPNet model outperforms existing methods and further provides interpretable biomarkers at regional activity, ROI-level connectivity, and subnetwork coordination levels.

2 Related Work

2.1 Spatial Brain Network Analysis

FC quantifies statistical dependencies between regional time series, naturally enabling graph-based representations of brain networks. Consequently, graph neural networks (GNNs) have become prevalent for modeling spatial dependencies. Early architectures like BrainGNN (Li et al., 2021) capture local interactions, later enhanced by adaptive region discrimination (Xu et al., 2024a) and information-theoretic pruning (Zheng et al., 2024). However, their limited receptive fields struggle with long-range dependencies. Graph transformers address this via global attention, as demonstrated by BrainNetTF (Kan et al., 2022) and ALTER (Yu et al., 2024), which integrate random-walk priors. LGC-SGT (Zhang et al., 2025) uses a spike-based graph transformer, enabling dual-perspective analysis of neural activity and connectivity. To capture the brain's multi-scale organization, recent hierarchical models, including HFBN (Zhang et al., 2024), DHGFormer (Xue et al., 2025), BioBGT (Peng et al., 2025), and BrainHGT (Ma et al., 2025), employ strategies such as bottom-up pooling, contrastive learning, or neuroanatomical priors to progressively aggregate nodes into hierarchical representations. Despite effectively modeling cross-scale functional integration, these approaches predominantly rely on undirected connectivity, overlooking the causal information flow within and across hierarchical levels.

2.2 Temporal Brain Network Analysis

Temporal brain network modeling aims to capture the dynamic evolution of brain activity over time from fMRI time series. While early approaches employed RNNs such as LSTM, recent work favors Transformer-based models for their ability to model long-range temporal dependencies via self-attention. MILC (Mahmood et al., 2020) combines CNNs and attention over sliding windows. DICE (Mahmood et al., 2022) and Glacier (Mahmood et al., 2023) estimate task-specific directed

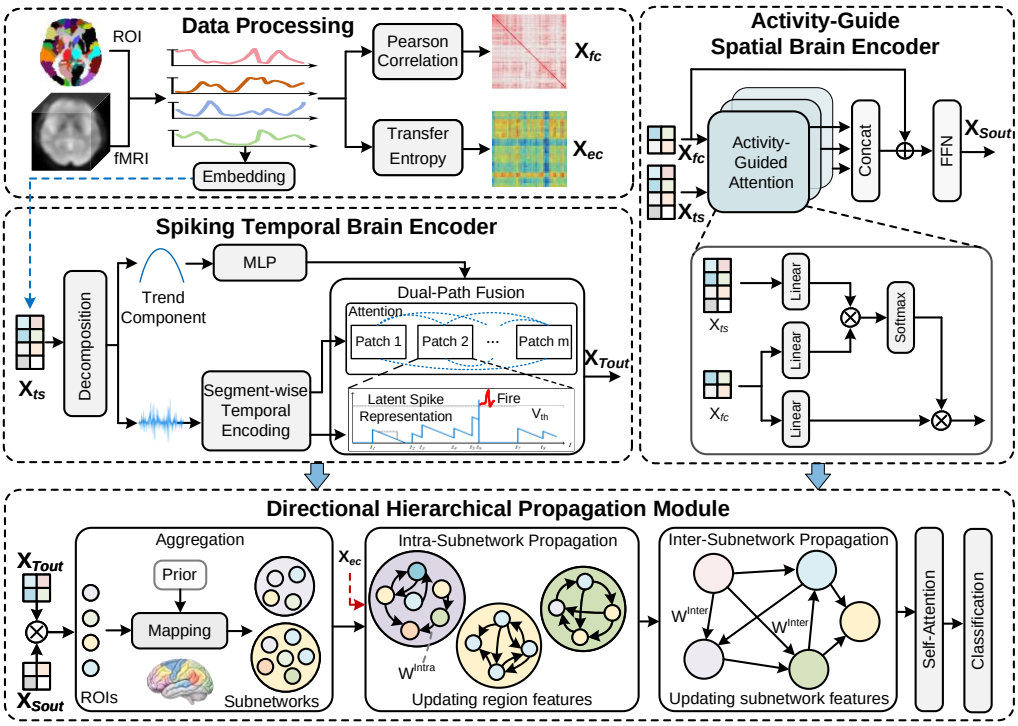


Figure 1: The overall framework of the proposed ST-DHPNet model.

connectivity matrices through spatiotemporal attention. BOLD Transformer (Bedel et al., 2023) adopts a hierarchical windowed design to enable local-to-global feature fusion and identify predictive time points. Despite progress in dynamic modeling, current approaches still tend to use generic sequential architectures that do not capture the brain's hierarchical organization.

3 METHODS

The overall architecture of the proposed model is illustrated in Figure 1. First, a spiking temporal brain encoder is introduced to capture the dynamic evolution of regional brain activity through stateful spiking computations on successive BOLD observations. Then, we design an activity-guided spatial brain encoder that leverages the activity states of individual brain regions to guide node representations. Finally, we develop a directional hierarchical propagation module that integrates ROI-level features into functional subnetwork communities. Building upon this hierarchy, we perform intra-subnetwork message passing guided by effective connectivity (EC) to refine the internal representations of each subnetwork. Subsequently, we conduct inter-subnetwork directed propagation based on the EC between subnetworks.

3.1 SPIKING TEMPORAL BRAIN ENCODER

We introduce a Spiking Temporal Brain Encoder (STBE) to model regional BOLD sequences as a nonlinear temporal dynamical system. Its internal membrane state integrates successive inputs, while thresholding and reset generate selective latent activations.

Decomposition. To isolate the intrinsic resting-state activity trends of the brain, we first decompose each regional BOLD time series into distinct components. Specifically, we separate the slowly varying baseline trend from transient functional fluctuations. This decomposition is implemented via a learnable 1D convolutional filter. We empirically set the stride to $S = 1$ and the kernel size to $K = 25$. The filter weights are initialized using a Gaussian-shaped profile followed by softmax normalization. Formally, we denote the learnable weights as $w \in \mathbb{R}^{1\times1\times K}$, parameterized by a

hyperparameter $\sigma = 1.0$, which controls the standard deviation of the Gaussian initialization, as

$$U \in \mathbb{R}^{1\times1\times K},$$
$$U[0,0,i] = \exp\left(-\frac{(i-K/2)^2}{2\sigma^2}\right),\ i = 1,2,\ldots,K,$$
$$\omega = \text{Softmax}(U, \text{dim} = -1). \tag{1}$$

To preserve sequence length during convolution, we pad $X_{ts} \in \mathbb{R}^{N\times T}$ with terminal values, yielding $X_{padded} \in \mathbb{R}^{N\times(T+K-1)}$. We then split it into N univariate series $X_i \in \mathbb{R}^{1\times(T+K-1)}$ ($i = 1,\ldots,N$) and apply a shared 1D convolutional kernel to extract the trend component $x_i \in \mathbb{R}^{1\times T}$. Concatenating all x_i gives $X_{trend} \in \mathbb{R}^{N\times T}$, and the residual component is obtained as $X_{res} = X_{ts} - X_{trend}$.

Segment-wise Temporal Encoding. Owing to the inherent smoothness of the trend component, we encode it using a simple MLP:

$$X_{Trend_{out}} = \text{MLP}(X_{trend}). \tag{2}$$

We partition the residual sequence into m consecutive, non-overlapping segments, each containing T_s BOLD frames, with $T = mT_s$. One frame corresponds to one computational SNN timestep, so each segment is processed sequentially over T_s steps. This convention defines a computational sequence index, not a physiological spike timescale. The number of segments is determined by $m = T/T_s$ rather than tuned independently. The structured input is

$$X_{embed} = \left[X_{res}^{(1)}, X_{res}^{(2)}, \ldots, X_{res}^{(m)}\right],$$
$$X_{res}^{(m)} \in \mathbb{R}^{N\times T_s}, X_{embed} \in \mathbb{R}^{m\times N\times T_s}. \tag{3}$$

The segmentation yields an intra-segment representation of consecutive observations and an inter-segment representation of dependencies across the m segments. The window width affects only temporal encoding: both FC and EC are estimated from the complete preprocessed BOLD sequence.

Fusion. The two paths are the trend branch and the residual dynamic branch. Within the dynamic branch, an SNN encodes each segment and self-attention models inter-segment dependencies. We use an integrate-and-fire update derived from the Leaky Integrate-and-Fire (LIF) formulation (Maass, 1997):

$$H[t] = \tau \cdot V[t-1] + X[t], \qquad S[t] = \Theta\big(H[t] - V_{th}\big), \tag{4}$$

$$V[t] = H[t] \cdot \big(1 - S[t]\big) + V_{reset} \cdot S[t], \tag{5}$$

where $X[t]$ denotes the input current at time step t, $H[t]$ represents the membrane potential after charging but before firing, $V[t]$ is the membrane potential after firing, $S[t]$ indicates the binary spike output, $\tau = 1.0$ is the discrete membrane retention factor, and $V_{th} = 1.0$ stands for the firing threshold. Θ is the Heaviside step function and $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ otherwise. Moreover, due to the non-differentiability of spiking signals, we adopt surrogate gradient-based direct training for our SNN. During backpropagation, we employ the derivative of the scaled hyperbolic tangent function $\varphi(x) = \frac{1}{2}\tanh\left(3\left(x - 0.5\right)\right) + \frac{1}{2}$ as the surrogate gradient to approximate the derivative of the Heaviside step function. The spiking layer converts the continuous signal into a spike train over timesteps, yielding an intra-segment representation:

$$X_{Intra}^{i} = \text{SN}(X_{res}^{i}), (i = 1,\ldots,m), \tag{6}$$

where SN represents the spiking layer. These segment-wise outputs $X_{Intra}^{i} \in \mathbb{R}^{N\times T_s}$ are then concatenated along the temporal dimension to reconstruct the full-length sequence representation $X_{Intra} \in \mathbb{R}^{N\times T}$. To capture long-range dependencies across segments, we further compute a compact inter-segment representation by applying self-attention to the embedded segment time series:

$$X_{Inter} = \text{Attn}(X_{embed}). \tag{7}$$

We fuse local latent dynamics with inter-segment context and retain the trend branch:

$$X_{T_{out}} = X_{Trend_{out}} + X_{Inter} \odot (X_{Intra}W), \tag{8}$$

where W projects the intra-segment features to the same shape as X_{Inter} and $X_{Trend_{out}}$, and $\odot$ denotes element-wise multiplication.

3.2 ACTIVITY-GUIDED SPATIAL BRAIN ENCODER

Existing attention-based methods for brain network modeling typically only rely on functional connectivity, decoupling the attention process from the actual neural activity. To bridge this gap, we propose an Activity-Guided Attention (AGA) mechanism that modulates FC edge weights based on the activity states of brain regions. Specifically, the strength of attention between two regions is adaptively adjusted according to their temporal BOLD signal patterns, ensuring that a stronger influence is assigned when there is high activation. This method preserves the original FC structure while enriching it with regional temporal activity patterns. This enables the model to prioritize biologically meaningful interactions during message passing.

The activity-guided spatial encoder takes as input both the BOLD time series X_{ts} and the latent functional connectivity matrix X_{fc} derived from it and outputs a set of global region-level representations:

$$X_{S_{out}} = \mathrm{Encoder_s}(X_{ts}, X_{fc}). \tag{9}$$

We adopt the transformer architecture as the backbone. In this design, the query is derived from the BOLD time series to capture regional temporal activity patterns, while the key and value are constructed from the FC matrix to encode inter-regional statistical dependencies. This enables the attention mechanism to dynamically modulate functional connections based on real-time neural activity.

The output of the i-th attention head is computed as

$$O^i = \mathrm{Softmax}(\frac{((W_Q X_{ts})^i)^T (W_K X_{fc})^i}{\sqrt{d_{head}}})(W_V X_{fc})^i, \tag{10}$$

where W_Q, W_K, W_V are learnable matrices and d_{head} is the dimension of each attention head. Finally, the multi-head outputs are concatenated and linearly projected to produce the spatial representation:

$$X_{S_{out}} = \mathrm{W}_O \left(\mathrm{Concat}(O^1, \ldots, O^i)\right). \tag{11}$$

3.3 DIRECTIONAL HIERARCHICAL PROPAGATION MODULE

While ROI-based representations provide region-level information, they often lack direct interpretability in terms of functional systems. In contrast, functional subnetworks group ROIs into biologically meaningful functional systems based on coherent activity patterns. Thus, ROIs capture fine-grained spatial detail, whereas functional subnetworks reflect system-level organization that aligns with known neurocognitive architectures. To bridge the gap between fine-grained regional representations and system-level brain organization, we introduce a directional hierarchical propagation module that leverages prior knowledge of canonical functional networks. Specifically, we leverage prior knowledge of functional partitioning to aggregate regions into subnetworks, learn intra-subnetwork local dependencies, and further update region representations accordingly.

Aggregation. To respect the prior functional organization while enabling data-driven ROI-to-subnetwork association, we introduce M learnable subnetwork prototypes $C \in \mathbb{R}^{M \times d}$ as initial anchors for subnetwork-level representations. Following BrainHGT (Ma et al., 2025), these prototypes are initialized with Xavier uniform initialization and orthogonalized via the Gram–Schmidt process. Additionally, we construct a spatial similarity prior matrix $D_{\text{prior}} \in \mathbb{R}^{N \times M}$, which encodes the anatomical correspondence between ROIs and Yeo-7 subnetworks (Yeo et al., 2011) to guide attention toward neurobiologically plausible assignments. For the default Yeo-7 prior, we include an additional community for cerebellar and subcortical regions (CB&SC), yielding $M = 8$ prototypes for the seven cortical networks and this additional community. Given the ROI-level spatiotemporal feature matrix $X_{st} = (W_T X_{T_{\text{out}}}) \odot X_{S_{\text{out}}} \in \mathbb{R}^{N \times d}$, we compute the prototype-to-ROI attention scores as

$$E = \frac{(CW_q)(X_{st}W_k)^\top}{\sqrt{d_{\text{head}}}} \in \mathbb{R}^{M \times N}, \tag{12}$$

where $W_q, W_k \in \mathbb{R}^{d \times d_{\text{head}}}$ are learnable projection matrices. To incorporate anatomical priors, we modulate these scores using the transposed prior matrix:

$$\tilde{E} = E \odot D_{\text{prior}}^\top, \tag{13}$$

where $\odot$ denotes element-wise multiplication. The continuous attention weights are then obtained via the Entmax transformation:

$$W_{\text{attn}} = \text{Entmax}(\tilde{E}) \in \mathbb{R}^{M \times N}. \tag{14}$$

The first-stage subnetwork representations are computed by aggregating ROI features with these attention weights:

$$\tilde{Z}_1 = W_{\text{attn}}(X_{st} W_1) \in \mathbb{R}^{M \times d}, \tag{15}$$

where $W_1 \in \mathbb{R}^{d \times d}$ is a learnable transformation. Separately, discrete ROI-to-subnetwork assignments are derived by applying the argmax operator along the prototype dimension:

$$a_i = \underset{k \in \{1,\ldots,M\}}{\text{argmax}} \; (W_{\text{attn}})_{ki}, \qquad \mathcal{G}_k = \{i : a_i = k\}, \tag{16}$$

where a_i denotes the subnetwork index assigned to the i-th ROI, and $\mathcal{G}_k$ is the set of ROIs assigned to the k-th subnetwork.

Intra-Subnetwork Propagation. We leverage a directed effective connectivity matrix $W^{EC} \in \mathbb{R}^{N \times N}$ to represent directed statistical dependencies among ROIs. We use the subnetwork partitioning from aggregation. For the k-th subnetwork ϱ_k, we extract its internal EC submatrix $W^{EC}_{\mathcal{G}_k} \in \mathbb{R}^{|\mathcal{G}_k| \times |\mathcal{G}_k|}$, where W_{ij} quantifies the estimated directed dependence from region i to region j. We perform directed aggregation in each subnetwork:

$$\tilde{x}_j = \sum_{i \in \mathcal{G}_k} W^{EC}_{ij} \, x_i, \quad \forall i, \forall j \in \mathcal{G}_k. \tag{17}$$

This operation uses the source-to-target orientation of estimated EC to weight regional messages. Finally, we apply mean pooling over the updated ROI features to obtain the second-stage subnetwork representation:

$$\tilde{Z}_2 = \frac{1}{|\mathcal{G}_k|} \sum_{j \in \mathcal{G}_k} \tilde{\mathbf{x}}_j. \tag{18}$$

Inter-Subnetwork Propagation. Furthermore, we construct an inter-subnetwork EC matrix $W^{Inter} \in \mathbb{R}^{M \times M}$, where each entry W^{Inter}_{pq} is computed as the average of all pairwise EC values from ROIs in source subnetwork $\mathcal{G}_p$ to those in target subnetwork $\mathcal{G}_q$.

$$W^{Inter}_{pq} = \frac{1}{|\mathcal{G}_p| \cdot |\mathcal{G}_q|} \sum_{i \in \mathcal{G}_p} \sum_{j \in \mathcal{G}_q} W_{ij}, \quad p, q = 1, \ldots, M. \tag{19}$$

Before propagation, $\tilde{Z}_1$ preserves the attention-aggregated subnetwork state, whereas $\tilde{Z}_2$ incorporates intra-subnetwork EC messages. Their sum retains both representations. With W^{Inter}_{pq} denoting source p to target q and subnetwork features stored as rows, propagation is

$$\tilde{Z}_q = \sum_p W^{Inter}_{pq} (\tilde{Z}_{1,p} + \tilde{Z}_{2,p}), \qquad \tilde{Z} = (W^{Inter})^\top (\tilde{Z}_1 + \tilde{Z}_2). \tag{20}$$

Finally, we apply a self-attention mechanism over the refined subnetwork representations to model interactions among functional subnetworks, followed by a feed-forward network to further enhance representational capacity.

4 EXPERIMENTS

Datasets, Data Preprocessing, and Implementation Details. We evaluate ST-DHPNet on three fMRI datasets. (1) ABIDE contains 516 patients with autism spectrum disorder (ASD) and 493 healthy controls (HCs). (2) ADNI contains 190 patients with mild cognitive impairment (MCI) and 200 HCs. (3) REST-meta-MDD contains 1240 patients with major depressive disorder (MDD) and 1101 HCs. Each dataset was randomly divided into 70% training, 10% validation, and 20% test sets. Experiments were repeated 10 times with different random seeds. The details of datasets, data preprocessing, and experimental implementation can be found in Appendix A.

Table 1: Performance comparison with existing methods on ABIDE, ADNI, and REST-meta-MDD datasets. We report the mean and standard deviation across 10 independent runs. **Bold** indicates the best results.

Method	Dataset: ABIDE			Dataset: ADNI			Dataset: REST-meta-MDD		
	ACC(%)↑	AUC(%)↑	SEN(%)↑	ACC(%)↑	AUC(%)↑	SEN(%)↑	ACC(%)↑	AUC(%)↑	SEN(%)↑
SVM	60.00±1.78	60.09±1.88	63.15±4.56	68.59±3.05	72.30±3.84	73.95±6.60	63.90±1.17	63.58±1.19	69.23±1.29
Random Forest	55.25±2.71	55.39±2.80	59.55±5.74	66.92±2.86	69.55±4.65	69.47±6.98	57.82±1.99	61.15±1.81	63.06±3.27
GAT	60.41±2.50	59.40±2.48	57.96±11.73	68.72±2.64	72.68±4.00	71.05±8.07	59.87±2.62	63.51±2.53	69.84±10.53
BrainGNN	57.74±4.40	62.10±3.38	39.96±24.04	68.97±4.01	73.20±4.95	66.32±7.61	67.61±2.22	70.92±2.47	69.76±9.80
IBGNN	59.85±2.78	60.67±3.27	64.56±13.88	67.82±2.78	72.80±4.40	64.21±7.65	64.65±1.80	70.13±1.72	71.03±8.42
BrainGB	63.12±3.24	66.32±2.97	65.10±12.01	66.15±2.00	71.99±3.91	67.89±6.94	58.70±1.37	62.99±1.15	68.21±6.84
BrainUSL	62.36±2.73	62.80±2.40	59.21±10.03	67.18±2.58	72.20±3.45	63.68±11.21	63.20±2.14	65.02±2.56	69.35±3.66
BrainIB	59.47±2.52	62.18±2.38	64.98±7.86	68.08±3.22	71.96±3.07	62.63±10.51	56.42±1.13	54.38±1.51	67.96±7.41
VanillaTF	64.33±2.70	72.97±2.02	62.60±11.30	69.23±2.67	78.49±2.56	76.47±6.32	63.96±2.18	70.73±1.15	64.25±5.80
BrainNetTF	67.98±3.63	75.63±2.74	70.31±15.69	70.51±2.77	78.28±2.71	71.05±6.82	65.03±2.32	69.46±1.47	65.73±4.99
GBT	67.34±2.37	75.67±3.37	65.47±13.83	69.23±2.82	78.04±2.07	73.69±6.58	64.39±1.20	69.70±1.64	70.16±3.28
RGTNet	63.35±3.06	68.42±3.50	63.83±10.0	67.95±2.86	74.93±3.69	66.53±5.39	60.87±2.19	62.61±2.84	65.32±6.98
Contrasformer	61.70±3.18	66.75±3.30	54.95±10.37	66.67±2.84	75.92±2.65	76.32±7.65	63.41±1.82	65.26±2.23	70.81±6.35
ALTER	68.87±1.89	75.80±2.63	69.98±6.56	70.51±2.76	76.45±2.74	63.15±6.53	66.79±2.25	71.01±1.64	61.13±4.87
BioBGT	63.97±1.74	62.60±1.77	60.91±6.01	71.79±2.36	71.45±3.58	67.09±7.18	59.83±3.22	59.79±2.47	60.48±4.91
DHGFormer	71.43±1.37	74.73±2.69	75.25±3.74	73.38±2.61	78.82±2.64	77.92±6.52	66.75±1.54	70.29±1.03	69.74±4.53
BrainHGT	71.33±1.91	76.65±1.45	72.40±3.96	73.59±2.88	79.07±2.46	73.95±6.91	67.46±1.09	71.52±0.81	69.03±4.34
SpikeGraphormer	69.56±2.12	75.27±0.76	72.55±7.26	68.21±3.38	79.12±1.14	66.32±9.62	64.69±1.11	69.90±1.03	67.10±9.96
SiGNN	67.09±1.05	75.17±1.72	56.78±5.47	69.23±3.34	79.09±2.76	67.89±6.09	64.90±1.49	70.25±0.87	66.94±1.82
LGC-SGT	69.46±1.71	77.06±1.78	73.66±8.54	71.54±3.48	78.76±0.42	67.89±8.87	65.33±1.69	69.51±0.70	67.18±6.68
ST-DHPNet	**73.20±2.06**	**78.02±2.21**	**78.51±4.97**	**74.87±2.20**	**79.71±2.34**	**80.52±6.98**	**68.02±2.01**	**71.67±1.70**	**71.81±4.17**

Baselines. We compare our model with four categories of baseline methods. The first category comprises conventional machine learning methods, including support vector machines (SVMs) and random forests. The second category comprises GNN-based methods, including GAT (Veličković et al., 2017), BrainGNN (Li et al., 2021), IBGNN (Cui et al., 2022b), BrainGB (Cui et al., 2022a), BrainUSL (Zhang et al., 2023), and BrainIB (Zheng et al., 2024). The third category comprises graph Transformer-based methods, including VanillaTF (Vaswani et al., 2017), BrainNetTF (Kan et al., 2022), GBT (Peng et al., 2024), ALTER (Yu et al., 2024), RGTNet (Wang et al., 2024), Contrasformer (Xu et al., 2024b), BioBGT (Peng et al., 2025), DHGFormer (Xue et al., 2025), and BrainHGT (Ma et al., 2025). The fourth category comprises three spiking neural network models: SpikeGraphormer (Sun et al., 2026), SiGNN (Chen et al., 2025), and LGC-SGT (Zhang et al., 2025).

Comparison with Existing Methods. Table 1 summarizes the experimental results on the ABIDE, ADNI, and REST-meta-MDD datasets, using accuracy (ACC), area under the receiver operating characteristic curve (AUC), and sensitivity (SEN). Our method achieves the highest mean ACC, AUC, and SEN among the compared methods on all three datasets. On the ABIDE dataset, the ACC of our model reaches 73.20%, which is 10.08% higher compared to the best GNN-based methods, 1.77% higher compared to the best Graph Transformer-based methods, and 3.64% higher compared to the best SNN-based methods. On the ADNI dataset, our model achieves a mean ACC of 74.87%, exceeding the best GNN-based method by 5.90%, the best Graph Transformer-based method by 1.28%, and the best SNN-based method by 3.33%. On the REST-meta-MDD dataset, ST-DHPNet achieves mean ACC, AUC, and SEN values of 68.02%, 71.67%, and 71.81%, respectively, with an ACC gain of 2.69% over the best SNN-based method. This performance gap underscores the advantage of integrating spiking temporal dynamics, activity-guided spatial encoding, and hierarchical effective connectivity in modeling the brain.

Ablation Study. We conduct comprehensive ablation studies to evaluate the contribution of each key component in our framework. The experimental design systematically isolates five key elements: (i) removing the spiking temporal brain encoder (w/o STBE), (ii) replacing our activity-guided attention with a standard attention mechanism that uses only FC (w/o AGA), (iii) removing series decomposition (w/o Decomposition), (iv) removing the hierarchical aggregation module (w/o Aggregation), and (v) removing the EC directional propagation (w/o EC). As shown in Table 2, removing any component leads to consistent performance degradation across all metrics. These results collectively confirm that each module plays a distinct and indispensable role in modeling brain pathology.

Analysis of Directional Hierarchical Propagation. We evaluate the robustness of ST-DHPNet to both functional-network priors and ROI parcellation schemes. First, with the CC200 parcellation fixed, we compare the Yeo-7 and Yeo-17 priors to examine the effect of subnetwork granularity. As shown in Table 3, both priors yield competitive performance across all three datasets, with

Table 2: Ablation studies of different components on three brain disorder classification tasks.

Component	Dataset: ABIDE			Dataset: ADNI			Dataset: REST-meta-MDD		
	ACC(%)↑	AUC(%)↑	SEN(%)↑	ACC(%)↑	AUC(%)↑	SEN(%)↑	ACC(%)↑	AUC(%)↑	SEN(%)↑
w/o AGA	72.71±2.38	77.58±2.69	75.06±7.19	74.61±1.97	79.26±2.63	78.42±5.61	67.27±1.15	71.60±1.13	71.40±4.47
w/o Decomposition	72.02±1.59	77.41±2.78	73.19±7.91	72.43±4.56	79.39±2.76	71.84±7.89	67.53±1.27	70.98±1.24	71.45±5.88
w/o STBE	72.36±1.98	77.23±2.17	75.49±7.06	73.59±2.31	78.08±2.42	80.26±6.78	67.23±0.93	71.51±1.08	70.12±6.63
w/o Aggregation	67.93±2.51	71.22±1.25	68.06±6.15	68.72±3.93	73.01±4.57	69.74±3.80	63.22±1.34	65.56±1.46	68.35±9.28
w/o EC	71.53±1.96	77.38±1.86	74.72±7.93	73.20±2.46	79.16±2.61	75.52±4.99	67.81±1.23	71.35±1.12	71.23±4.27
ST-DHPNet	**73.20±2.06**	**78.02±2.21**	**78.51±4.97**	**74.87±2.20**	**79.71±2.34**	**80.52±6.98**	**68.02±2.01**	**71.67±1.70**	**71.81±4.17**

Table 3: Model performance comparison using different functional-network priors with CC200 fixed.

Atlas	Dataset: ABIDE			Dataset: ADNI			Dataset: REST-meta-MDD		
	ACC(%)↑	AUC(%)↑	SEN(%)↑	ACC(%)↑	AUC(%)↑	SEN(%)↑	ACC(%)↑	AUC(%)↑	SEN(%)↑
w/o Atlas	70.69±2.54	77.03±2.83	70.96±7.07	70.51±3.29	78.39±2.51	74.75±10.67	66.71±2.12	70.64±2.09	70.06±7.31
Yeo 7-network	73.20±2.06	78.02±2.21	78.51±4.97	74.87±2.20	79.71±2.34	80.52±6.98	68.02±2.01	71.67±1.70	71.81±4.17
Yeo 17-network	71.58±2.81	77.01±2.63	78.47±6.50	74.36±3.29	78.78±3.28	78.16±8.49	67.44±1.63	71.60±1.66	71.13±4.98

Table 4: Model performance comparison using different ROI parcellations with the Yeo-7 prior fixed.

Atlas	Dataset: ABIDE			Dataset: ADNI			Dataset: REST-meta-MDD		
	ACC(%)↑	AUC(%)↑	SEN(%)↑	ACC(%)↑	AUC(%)↑	SEN(%)↑	ACC(%)↑	AUC(%)↑	SEN(%)↑
CC200	73.20±2.06	78.02±2.21	78.51±4.97	74.87±2.20	79.71±2.34	80.52±6.98	68.02±2.01	71.67±1.70	71.81±4.17
AAL116	72.37±2.34	77.75±2.48	76.07±11.46	73.78±3.36	81.24±2.04	82.35±5.58	68.70±1.64	71.60±1.90	70.34±9.08

Table 5: Model performance comparison using different propagation strategies.

Component	Dataset: ABIDE			Dataset: ADNI			Dataset: REST-meta-MDD		
	ACC(%)↑	AUC(%)↑	SEN(%)↑	ACC(%)↑	AUC(%)↑	SEN(%)↑	ACC(%)↑	AUC(%)↑	SEN(%)↑
Only Intra	67.63±1.20	71.66±1.33	70.36±4.76	74.31±1.97	79.36±2.32	76.35±6.18	67.32±1.33	71.52±1.21	70.47±4.91
Only Inter	72.32±2.53	77.41±2.84	75.83±6.31	74.24±3.69	79.29±2.91	78.66±6.43	67.35±1.31	71.37±1.45	70.83±4.72
ST-DHPNet	**73.20±2.06**	**78.02±2.21**	**78.51±4.97**	**74.87±2.20**	**79.71±2.34**	**80.52±6.98**	**68.02±2.01**	**71.67±1.70**	**71.81±4.17**

Yeo-7 performing slightly better on most metrics. Second, with the Yeo-7 prior fixed, we compare CC200 and AAL116 to assess sensitivity to ROI parcellation. All other experimental settings remain unchanged. Table 4 shows comparable performance under both atlases, with mean ACC differing by at most 1.09 percentage points across the three datasets. Together, these results support the robustness of ST-DHPNet to the tested choices of functional-network priors and ROI parcellations. Furthermore, to validate the necessity of the complete hierarchical propagation mechanism, we conduct an additional ablation study by isolating the two propagation pathways. As summarized in Table 5, while both isolated variants achieve reasonable performance, the full ST-DHPNet model consistently outperforms them across ACC, AUC, and SEN metrics on all datasets. These results confirm that the synergistic integration of both intra- and inter-subnetwork information flow is essential for capturing comprehensive hierarchical dependencies in brain networks, and that neither propagation pathway alone is sufficient to achieve optimal diagnostic performance.

5 Interpretability and Visualization

Thanks to its hierarchical modeling, our model provides a unique lens to investigate brain disorder mechanisms from three perspectives: regional neural activity, inter-regional connectivity, and functional subnetwork organization. In the following subsections, we conduct a systematic analysis of discriminative patterns at each hierarchical level using the ABIDE and ADNI datasets, offering interpretable neurobiological insights into how brain disorders manifest across distinct hierarchies.

Regional Neural Activity. We characterize regional latent activity by the mean model-generated spike rate across computational timesteps for each ROI. Patient–control differences are assessed using two-sample t-tests with false discovery rate (FDR) correction (Appendix C). Figure 2(a) shows a neuroanatomical schematic of the 10 ROIs with the largest intergroup spiking neural activity differences. The result reveals that the ASD group has lower latent activity in the highlighted frontal regions than HCs. This pattern is anatomically consistent with previously reported frontal abnormalities (Zhao et al., 2022).

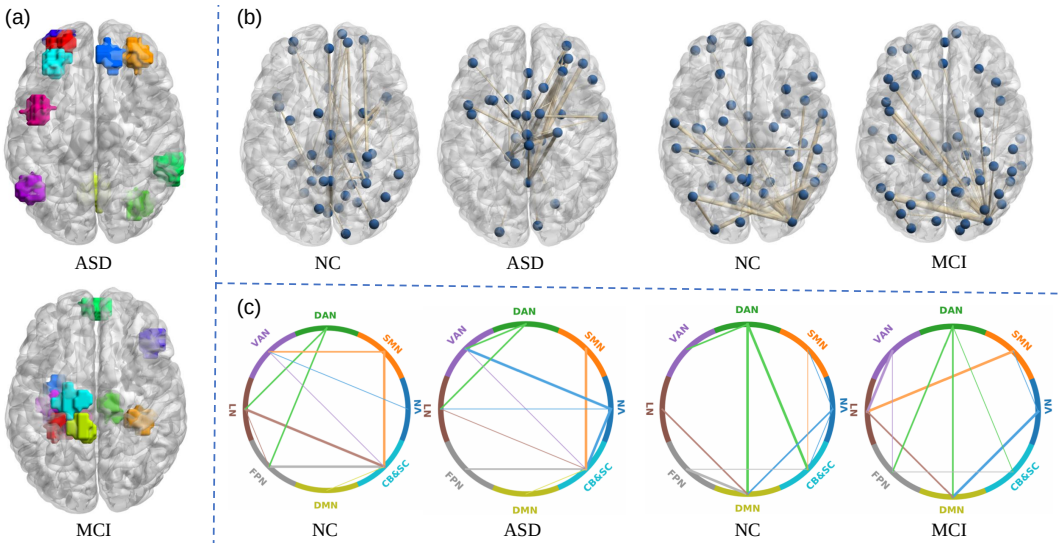


Figure 2: (a) Top 10 most discriminative regions in ASD and MCI compared to HCs. (b) Top 50 strongest ROI-level connections in HCs, ASD, and MCI. (c) Top 10 strongest subnetwork-level connections in HCs, ASD, and MCI.

Inter-Regional Connectivity. Next, we quantify ROI-level interactions using the learned ROI-to-ROI attention weights from the activity-guided spatial encoder. By comparing the 50 strongest functional connections in each group, we identify distinct connectivity patterns that differentiate individuals with ASD and MCI from their respective healthy controls, as illustrated in Figure 2(b). We observed that individuals with ASD exhibited reduced intra-network connectivity within the DMN compared to HCs, consistent with prior work reporting persistent DMN dysfunction in ASD (Lee et al., 2025). Given the critical role of the DMN in self-referential cognition, memory processing, and emotion regulation, this weakened connectivity may represent a key network-level substrate underlying the social cognitive impairments characteristic of ASD. Results demonstrate that our model identifies disorder-specific regional functional connectivity abnormalities relative to healthy controls.

Functional Subnetwork Organization. Finally, we quantify subnetwork-level interactions using attention weights between subnetworks after directed EC propagation. By aggregating regions into canonical functional subnetworks and propagating their representations using EC, we identify altered inter-subnetwork interactions, as shown in Figure 2(c). In ASD, we observe increased inter-subnetwork connectivity from the DAN to the VAN and from the SMN to the VAN, alongside decreased connectivity from the DAN and LN to the FPN. In contrast, MCI is characterized by elevated connectivity from the DAN and VAN to the FPN, as well as from the VAN to the LN, but reduced coupling from the FPN to the DMN and from the SMN to the LN. These results align well with observations in the literature (Qian et al., 2015; Padmanabhan et al., 2017).

6 Conclusion

We propose a spiking-temporal directed hierarchical propagation network that unifies spiking temporal dynamics and spatial hierarchy for brain disorder diagnosis, capturing multi-hierarchical information transmission from regional neural activity to subnetwork coordination. It combines spiking-based temporal encoding with activity-guided spatial encoding to capture regional neural activity and activity-dependent functional connectivity, while directional hierarchical propagation constructs a spatial hierarchy from individual regions to subnetworks, enabling cross-hierarchical information transmission. Experiments on three brain disorder datasets demonstrate superior diagnostic performance and identify interpretable biomarkers at three distinct levels: regional activity, inter-regional connectivity, and subnetwork coordination, providing a unified framework for understanding brain disorders across hierarchical levels.

AI USE STATEMENT

We used generative AI tools solely to improve the language, grammar, and clarity of the manuscript. These tools were not used to generate research ideas, develop methods, conduct experiments, analyze data, or produce results. All AI-assisted edits were reviewed by the authors to ensure that they accurately preserved the intended scientific meaning. The authors take full responsibility for the final content of this work.

REFERENCES

Hasan A Bedel, Irmak Sivgin, Onat Dalmaz, Salman UH Dar, and Tolga Çukur. Bolt: Fused window transformers for fmri time series analysis. *Medical image analysis*, 88:102841, 2023.

Dong Chen, Shuai Zheng, Muhao Xu, Zhenfeng Zhu, and Yao Zhao. Signn: A spike-induced graph neural network for dynamic graph representation learning. *Pattern Recognition*, 158:111026, 2025.

Nick Corriveau-Lecavalier, Jenna N Adams, Larissa Fischer, Eóin N Molloy, and Anne Maass. Cerebral hyperactivation across the alzheimer's disease pathological cascade. *Brain Communications*, 6(6):fcae376, 2024.

R Cameron Craddock, G Andrew James, Paul E Holtzheimer III, Xiaoping P Hu, and Helen S Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xuan Kan, Antonio Aodong Chen Gu, Joshua Lukemire, Liang Zhan, Lifang He, Ying Guo, and Carl Yang. Braingb: a benchmark for brain network analysis with graph neural networks. *IEEE transactions on medical imaging*, 42(2):493–506, 2022a.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. Interpretable graph neural networks for connectome-based brain disorder analysis. In *International conference on medical image computing and computer-assisted intervention*, pp. 375–385. Springer, 2022b.

Lulu Cui, Shu Li, Siman Wang, Xiafang Wu, Yingyu Liu, Weiyang Yu, Yijun Wang, Yong Tang, Maosheng Xia, and Baoman Li. Major depressive disorder: hypothesis, mechanism, prevention and treatment. *Signal transduction and targeted therapy*, 9(1):30, 2024.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. *Advances in Neural Information Processing Systems*, 35:25586–25599, 2022.

Jong-eun Lee, Sunghun Kim, Shinwon Park, Hyoungshin Choi, Bo-yong Park, and Hyunjin Park. Atypical maturation of the functional connectome hierarchy in autism. *Molecular autism*, 16(1): 21, 2025.

Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical Image Analysis*, 74:102233, 2021.

Yuanyuan Li, Yue Li, Qiang Wei, Tongjian Bai, Kai Wang, Jiaojian Wang, and Yanghua Tian. Mapping intrinsic functional network topological architecture in major depression disorder after electroconvulsive therapy. *Journal of Affective Disorders*, 311:103–109, 2022.

Rui Liu, Zhi-An Huang, Yao Hu, Zexuan Zhu, Ka-Chun Wong, and Kay Chen Tan. Spatial–temporal co-attention learning for diagnosis of mental disorders from resting-state fmri data. *IEEE transactions on neural networks and learning systems*, 35(8):10591–10605, 2023.

Jiajun Ma, Yongchao Zhang, Chao Zhang, Zhao Lv, and Shengbing Pei. Brainhgt: A hierarchical graph transformer for interpretable brain network analysis. *arXiv preprint arXiv:2511.17604*, 2025.

Wolfgang Maass. Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, 10(9):1659–1671, 1997.

Usman Mahmood, Md Mahfuzur Rahman, Alex Fedorov, Noah Lewis, Zening Fu, Vince D Calhoun, and Sergey M Plis. Whole milc: generalizing learned dynamics across tasks, datasets, and populations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 407–417. Springer, 2020.

Usman Mahmood, Zening Fu, Satrajit Ghosh, Vince Calhoun, and Sergey Plis. Through the looking glass: Deep interpretable dynamic directed connectivity in resting fmri. *NeuroImage*, 264:119737, 2022.

Usman Mahmood, Zening Fu, Vince Calhoun, and Sergey Plis. Glacier: glass-box transformer for interpretable dynamic neuroimaging. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

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.

Jorge J Palop and Lennart Mucke. Network abnormalities and interneuron dysfunction in alzheimer disease. *Nature Reviews Neuroscience*, 17(12):777–792, 2016.

Fabrizio Parente and Alfredo Colosimo. Modelling a multiplex brain network by local transfer entropy. *Scientific reports*, 11(1):15525, 2021.

Ciyuan Peng, Yuelong Huang, Qichao Dong, Shuo Yu, Feng Xia, Chengqi Zhang, and Yaochu Jin. Biologically plausible brain graph transformer. *arXiv preprint arXiv:2502.08958*, 2025.

Zhihao Peng, Zhibin He, Yu Jiang, Pengyu Wang, and Yixuan Yuan. Gbt: Geometric-oriented brain transformer for autism diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 142–152. Springer, 2024.

Adam R Pines, Bart Larsen, Zaixu Cui, Valerie J Sydnor, Maxwell A Bertolero, Azeez Adebimpe, Aaron F Alexander-Bloch, Christos Davatzikos, Damien A Fair, Ruben C Gur, et al. Dissociable multi-scale patterns of development in personalized brain networks. *Nature communications*, 13 (1):2647, 2022.

Shaowen Qian, Zhaoyan Zhang, Bo Li, and Gang Sun. Functional-structural degeneration in dorsal and ventral attention systems for alzheimer's disease, amnestic mild cognitive impairment. *Brain imaging and behavior*, 9(4):790–800, 2015.

Daniel A Rossignol and Richard E Frye. A review of research trends in physiological abnormalities in autism spectrum disorders: immune dysregulation, inflammation, oxidative stress, mitochondrial dysfunction and environmental toxicant exposures. *Molecular psychiatry*, 17(4):389–401, 2012.

Patricia Segura, Marco Pagani, Somer L Bishop, Phoebe Thomson, Stan Colcombe, Ting Xu, Zekiel Z Factor, Emily C Hector, So Hyun Kim, Michael V Lombardo, et al. Connectome-based symptom mapping and in silico related gene expression in children with autism and/or attention-deficit/hyperactivity disorder. *Molecular Psychiatry*, 31(1):282–295, 2026.

Olaf Sporns and Richard F Betzel. Modular brain networks. *Annual review of psychology*, 67(1): 613–640, 2016.

Sonja Sudimac, Vera Sale, and Simone Kühn. How nature nurtures: Amygdala activity decreases as the result of a one-hour walk in nature. *Molecular psychiatry*, 27(11):4446–4452, 2022.

Yundong Sun, Dongjie Zhu, Yansong Wang, Zhaoshuo Tian, Ning Cao, and Gregory O'Hare. Spikegraphormer: A high-performance graph transformer with spiking graph attention. *Information Sciences*, 755:123828, 2026.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Raul Vicente, Michael Wibral, Michael Lindner, and Gordon Pipa. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience*, 30(1):45–67, 2011.

Yibin Wang, Haixia Long, Tao Bo, and Jianwei Zheng. Residual graph transformer for autism spectrum disorder prediction. *Computer Methods and Programs in Biomedicine*, 247:108065, 2024.

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*, 2024a.

Jiaxing Xu, Kai He, Mengcheng Lan, Qingtian Bian, Wei Li, Tieying Li, Yiping Ke, and Miao Qiao. Contrasformer: a brain network contrastive transformer for neurodegenerative condition identification. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 2671–2681, 2024b.

Rundong Xue, Hao Hu, Zeyu Zhang, Xiangmin Han, Juan Wang, Yue Gao, and Shaoyi Du. Dhgformer: Dynamic hierarchical graph transformer for disorder brain disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 268–278. Springer, 2025.

BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 2011.

Shuo Yu, Shan Jin, Ming Li, Tabinda Sarwar, and Feng Xia. Long-range brain graph transformer. *Advances in Neural Information Processing Systems*, 37:24472–24495, 2024.

Geng Zhang, Jiangrong Shen, Kaizhong Zheng, Liangjun Chen, and Badong Chen. Local-global coupling spiking graph transformer for brain disorders diagnosis from two perspectives. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Jianjia Zhang, Luping Zhou, Lei Wang, Mengting Liu, and Dinggang Shen. Diffusion kernel attention network for brain disorder classification. *IEEE Transactions on Medical Imaging*, 41(10): 2814–2827, 2022.

Jianjia Zhang, Yunan Guo, Luping Zhou, Lei Wang, Weiwen Wu, and Dinggang Shen. Constructing hierarchical attentive functional brain networks for early ad diagnosis. *Medical Image Analysis*, 94:103137, 2024.

Pengshuai Zhang, Guangqi Wen, Peng Cao, Jinzhu Yang, Jinyu Zhang, Xizhe Zhang, Xinrong Zhu, Osmar R Zaiane, and Fei Wang. Brainusl: U nsupervised graph s tructure l earning for functional brain network analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 205–214. Springer, 2023.

Xiaoxin Zhao, Shuyi Zhu, Yang Cao, Peipei Cheng, Yuxiong Lin, Zhixin Sun, Wenqing Jiang, and Yasong Du. Abnormalities of gray matter volume and its correlation with clinical symptoms in adolescents with high-functioning autism spectrum disorder. *Neuropsychiatric Disease and Treatment*, 18:717, 2022.

Kaizhong Zheng, Shujian Yu, Baojuan Li, Robert Jenssen, and Badong Chen. Brainib: Interpretable brain network-based psychiatric diagnosis with graph information bottleneck. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

A Datasets, Data Preprocessing, and Implementation Details

A.1 Datasets

ABIDE Dataset. The ABIDE [1] dataset serves as a large-scale resource for ASD research, aggregating resting-state fMRI data from 17 international sites.

ADNI Dataset. The ADNI [2] dataset is a large-scale, multi-site repository designed to support research on Alzheimer's disease and its progression. It includes longitudinal neuroimaging, genetic, and clinical data collected from participants across multiple phases.

REST-meta-MDD Dataset. The REST-meta-MDD [3] dataset serves as a large-scale resource for MDD research, aggregating resting-state fMRI data from 25 research groups from 17 hospitals.

A.2 Data Preprocessing

Based on a predefined ROI atlas, the brain volume in each scan is parcellated into N ROIs. By averaging the BOLD signals of all voxels within each ROI, we obtain the regional activity time series $X_{ts} \in \mathbb{R}^{N\times T}$. Functional connectivity $X_{fc} \in \mathbb{R}^{N\times N}$ is then quantified by computing the Pearson correlation between every pair of time series, yielding a symmetric connectivity matrix that reflects the strength of statistical dependence between ROIs. In parallel, effective connectivity $X_{ec} \in \mathbb{R}^{N\times N}$ is estimated using transfer entropy to capture directed statistical dependencies between regional time series, using temporal-history parameter k.

BOLD Signals. We adopted a unified preprocessing workflow for all datasets using DPABI [4]. Specifically, the first 10 volumes of each fMRI time series were discarded to allow for signal stabilization. Slice-timing correction was applied to account for sequential slice acquisition, followed by head motion correction using a 6-parameter rigid-body transformation. Each subject's T1-weighted anatomical image was co-registered to the mean functional image and segmented into gray matter, white matter, and cerebrospinal fluid. We then employed DARTEL to estimate the nonlinear deformation field mapping native space to the MNI152 standard template. To mitigate low-frequency drifts in the BOLD signal, a linear trend regressor was included in the nuisance regression step. Finally, the CC200 functional brain atlas was used to parcellate the brain into $N = 200$ ROIs, and the BOLD time series for each ROI was extracted by averaging the signals of all voxels within the corresponding parcel.

Functional Connectivity. FC is commonly quantified as the Pearson correlation coefficient between the BOLD time series of pairs of brain regions. Let $\mathbf{x}_i = [x_i(1), x_i(2), \ldots, x_i(T)]^\top$ and $\mathbf{x}_j = [x_j(1), x_j(2), \ldots, x_j(T)]^\top$ denote the fMRI time series of region i and region j, respectively, each consisting of T time points. The functional connectivity C_{ij} between regions i and j is computed as:

$$C_{ij} = \frac{\sum_{t=1}^{T} \big(x_i(t) - \bar{x}_i\big)\big(x_j(t) - \bar{x}_j\big)}{\sqrt{\sum_{t=1}^{T} \big(x_i(t) - \bar{x}_i\big)^2}\sqrt{\sum_{t=1}^{T} \big(x_j(t) - \bar{x}_j\big)^2}}, \tag{21}$$

where $\bar{x}_i$ and $\bar{x}_j$ are the temporal means of $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively. The resulting FC matrix is symmetric, with values in $[-1, 1]$, where higher positive values indicate stronger synchronized activity between regions.

Effective Connectivity. Here, EC denotes estimated directed statistical dependence between regional BOLD sequences. In this work, we quantify EC using transfer entropy (TE) (Vicente et al., 2011), an information-theoretic measure that captures the amount of directed information flow from a source time series to a target time series beyond what is predictable from the target's own past. Let $\mathbf{x}_i$ and $\mathbf{x}_j$ denote the BOLD time series of regions i and j, respectively. The transfer entropy from region i to region j is defined as:

[1] http://preprocessed-connectomes-project.org/abide/

[2] https://adni.loni.usc.edu/

[3] https://rfmri.org/maps

[4] http://www.rfmri.org/dpabi

$$\mathrm{TE}_{i\to j} = \sum \left[p\big(x_j(t+1), x_j^{(k)}(t), x_i^{(l)}(t)\big) \times \log_2 \frac{p\big(x_j(t+1) \mid x_j^{(k)}(t), x_i^{(l)}(t)\big)}{p\big(x_j(t+1) \mid x_j^{(k)}(t)\big)} \right], \tag{22}$$

where $x_j^{(k)}(t) = [x_j(t), x_j(t-1), \ldots, x_j(t-k+1)]$ is the k-dimensional embedding of the past states of region j, $x_i^{(l)}(t) = [x_i(t), x_i(t-1), \ldots, x_i(t-l+1)]$ is the l-dimensional embedding of the past states of region i, and $p(\cdot)$ denotes the joint or conditional probability distribution estimated from the data.

A higher $\mathrm{TE}_{i\to j}$ indicates greater predictive information from region i about region j conditional on the included target history. The resulting effective connectivity matrix $\mathbf{E} \in \mathbb{R}^{N\times N}$ is generally asymmetric, reflecting the directional nature of neural interactions.

Both FC and TE are estimated once per subject from the complete preprocessed BOLD sequence, not from the short SNN segments. The initial TE setting follows prior work (Parente & Colosimo, 2021), and candidate settings were evaluated on the validation set. The default is $k = 1$; Appendix B.2 examines $k \in \{1, 2, 3, 4, 5\}$. This classification sensitivity analysis is not a direct test–retest assessment of TE estimates.

A.3 Implementation Details

The detailed hyperparameter settings for our model on the three datasets are summarized in Table 6. All experiments were performed on a single NVIDIA RTX 3090 GPU.

Table 6: Training hyperparameters for the proposed model.

Hyperparameter	**Value**	**Hyperparameter**	**Value**
Optimization	Adam	Learning rate	1e-4
Window width T_s	4	Weight decay	1e-4
Hidden dimensions	256	FFN hidden dimensions	1024
Assignment categories	8	Batch size	32
EC time lag coefficient	1	Attention Layers	1
Attention heads	8	Dropout rate	0.1
Readout method	mean	Epochs	100

B Additional Experiments and Analysis

B.1 Analysis of Spiking Temporal Dynamics

To rigorously evaluate the contribution of our spiking temporal dynamics module, we conducted a comprehensive comparison against widely adopted temporal modeling architectures, including recurrent networks (LSTM, GRU), temporal convolutional networks (TCN), and Transformer networks. All variants were integrated into the identical backbone under strictly matched training protocols to ensure a fair assessment of their temporal representation capabilities. As summarized in Table 7, the spiking encoder achieved the highest mean ACC, AUC, and SEN on ABIDE.

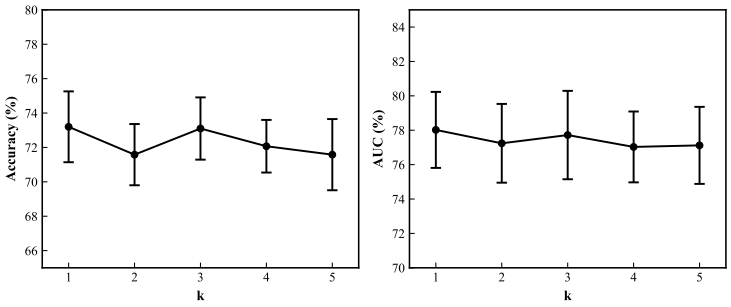
Table 7: Performance comparison with other widely adopted temporal modeling methods on the ABIDE dataset. **Bold** indicates the best results.

Datasets	**Component**	**ACC(%)** ↑	**AUC(%)** ↑	**SEN(%)** ↑
ABIDE	LSTM	70.89±1.47	75.90±2.19	68.37±8.48
	GRU	71.58±1.68	76.59±1.45	74.74±7.28
	TCN	70.20±2.30	75.10±3.25	72.90±9.79
	Transformer	72.56±1.72	77.06±2.19	75.10±8.72
	Spiking Neurons	**73.20±2.06**	**78.02±2.21**	**78.51±4.97**

B.2 Robustness Analysis of Directional Propagation

To evaluate the robustness of directional propagation, we constructed EC matrices using transfer entropy with different target-history lengths $k \in \{1, 2, 3, 4, 5\}$ in equation 22. These matrices were then used to guide hierarchical propagation. As shown in Figure 3, classification performance on ABIDE remains consistent across the tested settings, supporting the robustness of ST-DHPNet to the choice of target-history length for EC estimation.

Figure 3: Performance of directional propagation with varying target-history lengths (k) for transfer entropy estimation on ABIDE.



B.3 Sensitivity to Temporal Segmentation

We varied the segment width and corresponding number of SNN steps jointly over $T_s \in \{2, 4, 10\}$ on ABIDE and ADNI. For a fixed sequence length, $m = T/T_s$. As shown in Table 8, a segment width of four frames achieves the highest mean ACC and SEN on both datasets among the tested settings, supporting its use as the default configuration.

Table 8: Sensitivity to segment width T_s, which also sets the number of SNN computational steps per segment.

Dataset	T_s	ACC (%)	AUC (%)	SEN (%)
ABIDE	2	72.86±2.43	78.91±2.71	77.27±7.98
	4	73.20±2.06	78.02±2.21	78.51±4.97
	10	72.32±2.02	78.85±2.15	77.07±5.46
ADNI	2	73.97±2.75	80.41±3.18	80.00±9.21
	4	74.87±2.20	79.71±2.34	80.52±6.98
	10	73.62±2.86	78.87±3.29	79.42±6.91

B.4 Statistical Comparisons

To assess whether the performance improvements were statistically reliable, we compared ST-DHPNet with BrainHGT and DHGFormer, both of which exhibited competitive overall performance. We conducted two-sided paired t-tests across ten matched runs using identical random seeds and data partitions. At a significance threshold of $p < 0.05$, ST-DHPNet achieved statistically significant improvements in the majority of comparisons.

B.5 Held-Out NYU Site Evaluation

ABIDE contains data collected from multiple independent acquisition sites. We therefore conducted a strict held-out cross-site validation. Specifically, the largest site, NYU, was used exclusively for testing, while model training, validation, and hyperparameter selection were performed using the remaining sites. This setting evaluates whether the model generalizes to an unseen acquisition site. As shown in Table 10, ST-DHPNet achieved the best performance among the compared methods.

Table 9: Reported two-sided paired t-test p-values comparing ST-DHPNet with two competitive baselines over ten matched runs.

Comparator	Metric	ABIDE	ADNI	REST-meta-MDD
BrainHGT	ACC	0.0001	0.0132	0.1516
	AUC	0.0032	0.0941	0.6457
	SEN	0.0001	0.0001	0.0073
DHGFormer	ACC	0.0006	0.0035	0.0024
	AUC	0.0001	0.0325	0.0009
	SEN	0.0015	0.0271	0.0091

To evaluate biomarker reproducibility, we used the NYU site as an independent held-out cohort and analyzed the brain regions showing the largest differences in spiking neural activity between healthy controls and patients. As reported in Table 11, the mean spike count for each brain region was compared between the two groups using two-sample t-tests with FDR correction, with the significance threshold set at $p < 0.05$. The five regions exhibiting the largest group differences were lIFG.2, rMFG.7, lMFG.3, rSFG.2, and lFP.3. These regions showed substantial anatomical overlap with the biomarkers identified in our primary multi-site analysis. Moreover, frontal regions independently reported in previous neuroimaging studies were anatomically consistent with those identified in our analysis (Segura et al., 2026). These converging results support the cross-site reproducibility and neurobiological relevance of the identified biomarkers.

Table 10: Classification on the NYU site.

Model	ACC (%)	AUC (%)	SEN (%)
DHGFormer	64.68±2.92	69.50±1.98	68.40±6.47
BrainHGT	70.11±1.46	73.42±4.42	73.40±7.72
ST-DHPNet	72.17±2.21	79.44±1.87	76.80±5.49

Table 11: Regional latent spike counts at the NYU site. All tabulated p-values are FDR-adjusted.

Group / statistic	lIFG.2	rMFG.7	lMFG.3	rSFG.2	lFP.3
Healthy controls	26.73	25.67	25.54	25.70	26.79
ASD	22.56	22.53	22.64	23.25	24.71
FDR-adjusted p-value	0.0032	0.0069	0.0256	0.0202	0.0308

C Brain Region Assignments

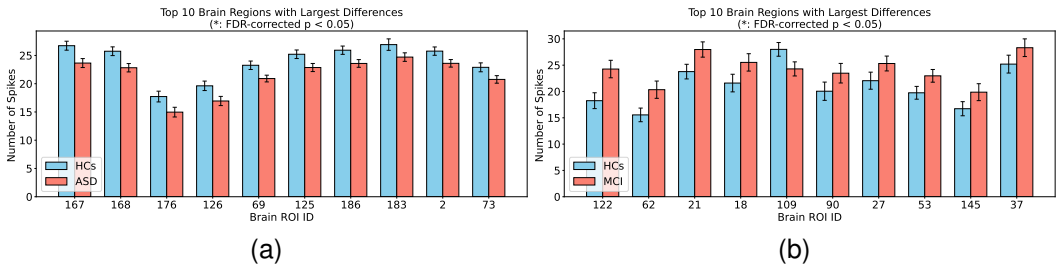


Figure 4: The top 10 most discriminative regions are shown. A two-sample t-test with FDR correction was used, and the significance threshold was $p < 0.05$. (a) Differential regional spike activity representations in ASD compared to HCs. (b) Differential regional spike activity representations in MCI compared to HCs.

The preprocessing of fMRI data follows the standardized pipeline. Both the ABIDE and ADNI datasets employ the Craddock 200 (CC200) atlas (Craddock et al., 2012) for brain parcellation, which partitions the brain into 200 ROIs. In this work, we partition the ROIs into eight functional systems based on an extended version of the Yeo 7-network atlas. The seven canonical cortical networks are default mode network (DMN), frontoparietal network (FPN), limbic network (LN), ventral attention network (VAN), dorsal attention network (DAN), somatomotor network (SMN), and visual network (VN), with an additional category for ROIs outside these cortical assignments. The anatomical listing in Table 12 labels this category cerebellar and subcortical structures (CB&SC). Table 12 lists the ROI assignments to eight communities on ABIDE: seven cortical functional networks corresponding to the Yeo-7 prior and an additional community comprising cerebellar and subcortical regions (CB&SC). Each ROI is identified by its name and index in the CC200 atlas.

As shown in Figure 4, we performed statistical comparisons of mean model-generated spike rates across all brain regions between individuals with ASD or MCI and HCs to identify key discriminative loci that may serve as focal biomarkers of pathology.

Table 12: Functional module and brain region assignments.

Functional Module	Brain Region (Name & Index)
DMN	lANG.1 (2), lACC.1 (5), lMTG.1 (11), rANG.2 (14), rACC.1 (22), rPCC (46), rMTG.2 (49), lACC.2 (51), rFOG.1 (53), rPRECU.3 (58), lMFG.2 (61), lMTG.2 (72), lFOG.2 (74), lANG.3 (82), rMEDFG.2 (91), lITG.2 (101), lFP.2 (104), rMFG.4 (106), rMTG.4 (107), rFP.3 (109), lMTG.3 (117), lSTG.2 (129), lSFG.3 (133), rFP.5 (139), rMTG.5 (140), lIFG.1 (141), rIFG.2 (144), rSTG.3 (153), rANG.3 (166), lIFG.2 (167), lSFG.4 (173), lPRECU.3 (174), rTP.3 (181), rSFG.2 (186), rSFG.3 (187), lSFG.5 (191), rSFG.4 (193), lINS.5 (196)
FPN	rANG.1 (7), rFP.1 (12), lMFG.1 (23), rMFG.1 (25), lpreC.1 (34), rMFG.2 (38), rMTG.1 (39), rACC.2 (40), lFP.1 (42), lANG.2 (56), rFP.2 (75), lSFG.2 (95), lITG.1 (99), rMFG.5 (113), rpreC.4 (115), rFP.4 (124), rMFG.6 (127), lMEDFG.1 (149), lMFG.4 (151), rIFG.3 (164), lIFG.3 (169), lFP.3 (183)
LN	lPARAH.1 (27), rTP.1 (32), lTP.1 (43), lFOG.1 (57), rFOG.2 (71), lTP.2 (78), rFUS.2 (87), rTP.2 (110), lFOG.3 (112), lPARAH.3 (145), rPARAH.2 (155), rACC.4 (160), rFUS.5 (198)
VAN	lINS.1 (4), rPRECU.2 (6), rMEDFG.1 (13), lINS.2 (20), rCING (29), lSUPRAM (33), rACC.3 (55), rINS.2 (59), rMTG.3 (69), lPCC (76), lACC.3 (79), rSTG.2 (83), lpostC.5 (116), rIFG.1 (119), lMFG.3 (125), rIPG (128), rINS.3 (137), rSFG.1 (161), lpreC.4 (165), rMFG.7 (168), rpostC.4 (180), lSMA (182), lINS.4 (184)
DAN	lIOG.2 (16), rpreC.1 (17), rSPL.1 (31), lSFG.1 (50), lFUS.1 (63), rMFG.3 (64), lpreC.2 (73), rSUPRAM (93), rITG (100), lSOG.2 (114), rSPG (132), lSOG.3 (136), rIOG.4 (143), lSPL.1 (156), rPRECU.4 (163), lSPL.2 (171), lSPL.3 (188), lPRECU.4 (197)
SMN	lpostC.1 (8), lpostC.2 (21), lSTG.1 (24), rpostC.1 (28), rINS.1 (35), rpostC.2 (60), rSPL.2 (65), rSTG.1 (66), lpostC.3 (88), lpreC.3 (90), lpostC.4 (96), rpreC.2 (98), rpreC.3 (111), lINS.3 (121), rparaC (123), rpreC.5 (134), lSTG.3 (146), lpostC.6 (154), rpostC.3 (157), rTTG (185), lTTG (200)
VN	lIOG.1 (1), rPRECU.1 (3), lPRECU.1 (19), rIOG.1 (26), rOP.1 (44), rPARAH.1 (48), lCUN (54), rFUS.1 (62), lLING.1 (70), rCUN (81), rIOG.2 (85), rLING.1 (89), lSOG.1 (97), rIOG.3 (102), rLING.2 (105), lLING.2 (108), lPARAH.2 (122), lIOG.3 (131), rFUS.3 (138), rMOG (142), lPRECU.2 (147), lIOG.4 (150), lOP (158), lFUS.2 (159), rSOG (170), rFUS.4 (172), rIOG.5 (175), lLING.3 (177), lLING.4 (179), lFUS.3 (189), rLING.3 (195)
CB&SC	lCEREB.1 (9), rCEREB.1 (10), rPUT.1 (15), rTHAL.1 (18), rBSTEM.1 (30), lCEREB.2 (36), lTHAL.1 (37), rCEREB.2 (41), lTHAL.2 (45), lCAUD.1 (47), lBSTEM.1 (52), lPUT (67), rMIDB.1 (68), rCEREB.3 (77), lCEREB.3 (80), lCAUD.2 (84), rBSTEM.2 (86), lAMY (92), rCAUD.1 (94), lCEREB.4 (103), rCEREB.4 (118), lCEREB.5 (120), rCEREB.5 (126), rMIDB.2 (130), rCAUD.2 (135), rTHAL.2 (148), rCEREB.6 (152), lCEREB.6 (162), rCEREB.7 (176), rPUT.2 (178), rBSTEM.3 (190), lCEREB.7 (192), rMIDB.3 (194), lBSTEM.2 (199)