HYPEREVO: HISTORY-AWARE DISCRETE HYPERGRAPH TOPOLOGY EVOLUTION FOR ENHANCED BRAIN DISEASE CLASSIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Functional brain networks exhibit complex high-order interactions and temporally evolving patterns, and their abnormal reorganization may provide important information for brain disease classification. However, existing methods often rely on predefined or fixed topologies, which may fail to capture disease-associated reorganization of dynamic node–hyperedge relationships across temporal states. Moreover, historical functional states are rarely exploited to guide topology evolution, limiting the ability to characterize the temporal evolution of functional brain networks. To address these problems, we propose **HyperEvo**, a history-aware discrete **hyper**graph topology **Evo**lution framework that improves brain disease classification by modeling dynamic functional brain networks. First, we construct window-level sparse hypergraphs from resting-state functional magnetic resonance imaging (rs-fMRI) signals and learn high-order functional representations through hypergraph convolution. Then, we aggregate preceding-window information through a history-aware topology evidence mechanism and perform budget-constrained discrete topology evolution by explicitly adding and deleting node–hyperedge memberships. Finally, we employ a multi-window kernel attention mechanism to integrate functional representations across temporal states and capture discriminative spatiotemporal patterns in dynamic functional brain networks. Extensive experiments on the ADNI and PPMI datasets demonstrate that HyperEvo achieves superior performance across multiple brain disease classification tasks. Further analyses show that the learned dynamic hypergraphs highlight interpretable brain regions and high-order functional connectivity patterns associated with disease-related functional alterations.

1 INTRODUCTION

Recent advances at the intersection of neuroscience and artificial intelligence have provided new technical routes to characterize the organizational principles of brain function and aid in the classification of brain diseases. Functional brain networks describe interaction patterns among brain regions and provide an important basis for analyzing abnormal brain states and identifying disease-related features (Ji & Zhang, 2022; Song et al., 2023). Studies have shown that brain diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD) are commonly accompanied by abnormal functional connectivity and altered spatiotemporal organization (Szewczyk-Krolikowski et al., 2014; Zhou et al., 2025). Resting-state functional magnetic resonance imaging (rs-fMRI) characterizes spontaneous functional interactions between brain regions through blood oxygen level-dependent (BOLD) signal fluctuations and has been widely used for functional brain network analysis and brain disease diagnosis (Wen et al., 2024).

Functional brain networks exhibit complex high-order interactions among multiple brain regions. Conventional graph representations mainly describe pairwise relationships and are therefore insufficient for modeling coordinated interactions involving multiple regions simultaneously (Eavani et al., 2015). Hypergraph models provide a natural framework for capturing such high-order functional dependencies, where hyperedges can simultaneously connect multiple regions of interest (ROIs) and represent complex functional coordination patterns (Jie et al., 2016; Zu et al., 2016; Gu et al., 2017). However, existing functional-brain hypergraph methods typically construct node–hyperedge mem-

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.

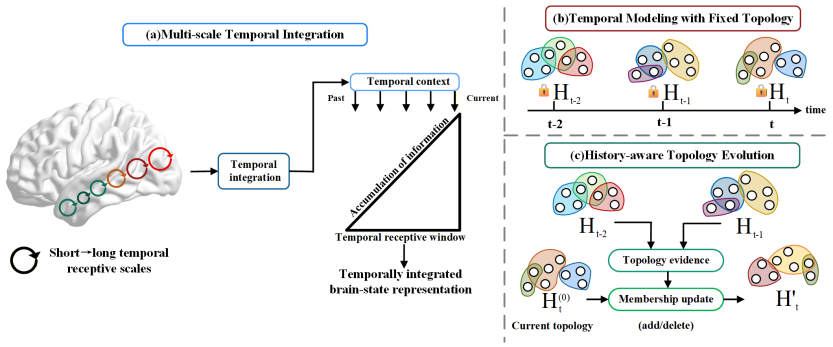

Figure 1: Illustration of spatio-temporal hypergraph modeling and history-aware topology evolution. (a) Multi-scale temporal integration for brain-state representation learning. (b) Spatio-temporal hypergraph modeling with preserved topology across time. (c) History-aware topology evolution guided by historical topology evidence.

berships based on predefined similarity metrics, clustering strategies, or sparse representations, and preserve these relationships unchanged during subsequent feature learning (Jie et al., 2016; Wang et al., 2023b; Ma et al., 2026). Although recent studies introduce adaptive hyperedge weights or attention mechanisms to adjust the importance of existing relations (Wang et al., 2023b; Li et al., 2026b), they mainly modify relation contributions rather than explicitly evolving node–hyperedge memberships.

Beyond structural complexity, brain activity also exhibits strong temporal continuity and multi-timescale integration characteristics. Brain regions at different hierarchical levels possess distinct temporal receptive windows and can integrate information from preceding states to form temporally contextualized representations (Chang et al., 2022). As illustrated in Fig. 2(a), multi-scale temporal integration aggregates information across different temporal contexts to construct brain-state representations with historical dependencies. Previous studies have demonstrated that current functional brain states are continuously influenced by preceding network states (Li et al., 2026b). Existing dynamic brain network methods mainly exploit temporal dependencies through feature-level aggregation mechanisms, such as temporal convolution, attention mechanisms, or state propagation (Kim et al., 2021; Ma et al., 2026; Li et al., 2026b). However, as shown in Fig. 2(b), existing dynamic hypergraph approaches generally construct hypergraphs independently for different temporal windows and perform temporal information interaction through feature propagation, while keeping node–hyperedge memberships unchanged. Since each temporal window only provides a local observation of brain activity, fixed hypergraph structures may fail to characterize the continuous reorganization of functional relationships over time.

To address these challenges, we propose HyperEvo, a history-aware discrete hypergraph topology evolution framework for dynamic brain network modeling and brain disease diagnosis. As illustrated in Fig. 2(c), HyperEvo exploits historical topology evidence to guide the adaptive evolution of current hypergraph structures. Specifically, HyperEvo first constructs window-level sparse hypergraphs from rs-fMRI signals to represent high-order functional interactions among brain regions. It then introduces a history-aware topology evidence mechanism that aggregates previous brain states through neighboring-window gating and multi-window historical attention, providing structural evidence for current topology adjustment. Based on this evidence, we design a budget-constrained discrete topology evolution strategy that adaptively updates node–hyperedge memberships through learnable additions and deletions. Finally, current-window features are re-encoded on the evolved hypergraph, and multi-window kernel attention integrates representations from different temporal states for cognitive-state classification.

To summarize, our contributions are three-fold:

- We identify the limitation of fixed hypergraph topology modeling in dynamic brain networks and propose HyperEvo, a history-aware discrete hypergraph topology evolution framework that enables adaptive topology updating for brain disease diagnosis.

- We introduce a history-aware topology evidence mechanism and a budget-constrained discrete evolution strategy, enabling historical brain states to guide node–hyperedge membership addition and deletion.
- We validate HyperEvo on ADNI and PPMI datasets, demonstrating improved classification performance and interpretable high-order functional connectivity patterns for brain disease diagnosis.

2 Related Work

Dynamic Functional Brain Network Analysis. Functional brain network analysis has evolved from static connectivity modeling to dynamic approaches that characterize temporal variations in brain functional organization. Traditional static networks estimate connectivity over the entire fMRI sequence but assume relatively stable functional relationships, which may overlook transient brain-state changes (Teng et al., 2024; Han et al., 2025). Dynamic functional brain network methods divide fMRI signals into multiple temporal windows and model evolving connectivity patterns across different brain states (Wang et al., 2021; Kim et al., 2021; Zhu et al., 2022; Liu et al., 2023). Recent approaches further incorporate temporal convolution, attention mechanisms, and state propagation strategies to capture cross-window dependencies and spatio-temporal representations (Zhu et al., 2024; Li et al., 2026b).

Hypergraph-based Brain Network Modeling. Conventional graph representations mainly model pairwise interactions and are insufficient for characterizing high-order coordination among multiple brain regions. Hypergraph models address this limitation by allowing hyperedges to connect multiple regions simultaneously and have been widely applied to functional brain network analysis (Eavani et al., 2015; Jie et al., 2016; Zu et al., 2016; Gu et al., 2017). Existing studies construct brain hypergraphs through similarity metrics or sparse representation and further improve representation learning using hyperedge weighting, attention mechanisms, or learnable structures (Wee et al., 2014; Wang et al., 2023b; Li et al., 2026a; Hu et al., 2026). However, most approaches preserve node–hyperedge memberships after construction and mainly optimize relation importance rather than explicitly modeling topology evolution.

Spatiotemporal Dynamic Brain Network Modeling. To jointly model temporal dynamics and high-order interactions, recent studies have explored spatiotemporal hypergraph learning for brain network analysis. These methods employ temporal similarity, attention mechanisms, window-specific hypergraphs, and spatiotemporal aggregation strategies to integrate functional information across multiple temporal states (Li et al., 2026a; Ma et al., 2026; Hu et al., 2026). Some studies further exploit cross-window dependencies through temporal propagation or feature fusion to capture the temporal evolution of brain states (Zhu et al., 2024; Li et al., 2025b;a; 2026b). Nevertheless, existing methods mainly exploit historical states at the representation level, while their influence on current hypergraph topology adaptation remains insufficiently explored.

3 Method

We propose HyperEvo, a history-aware discrete hypergraph topology evolution framework for dynamic brain network modeling and brain disease diagnosis.First, we construct the **Hypergraph Encoding** (§3.1) to construct window-level sparse coefficient hypergraphs from rs-fMRI signals and learn high-order functional representations among brain regions. Then, we propose the **History-aware Topology Evolution** (§3.2), which exploits historical brain states as topology evidence to adaptively update node–hyperedge memberships. Finally, we develop the **Kernel Attention Fusion** (§3.3), which integrates functional representations across different temporal windows for brain disease diagnosis.

3.1 Hypergraph Encoding

Hypergraph Initialization. To characterize high-order functional interactions among brain regions, we first construct window-specific sparse hypergraphs from rs-fMRI signals. Given an rs-fMRI sequence $X \in \mathbb{R}^{T\times N}$, we divide it into W non-overlapping temporal windows $\{X_t\}_{t=1}^{W}$, where $X_t \in \mathbb{R}^{L\times N}$ denotes the BOLD signals of the t-th window. For each temporal window, we construct

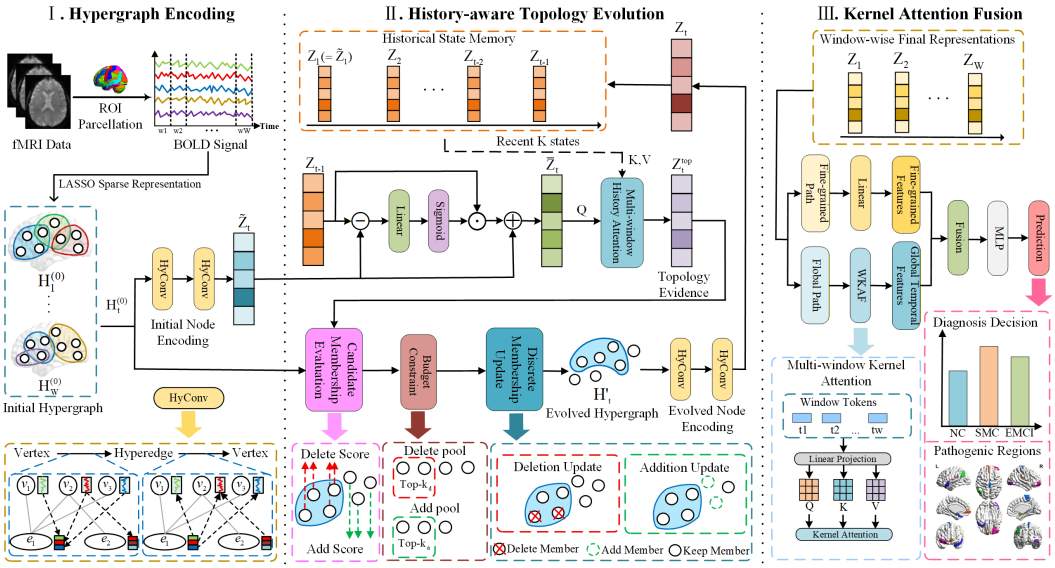

Figure 2: Overview of the HyperEvo framework. (I) Hypergraph Encoding constructs window-level hypergraphs to capture high-order functional interactions from rs-fMRI signals. (II) History-aware Topology Evolution exploits historical topology evidence to adaptively refine node–hyperedge memberships through discrete structural updates. (III) Kernel Attention Fusion integrates complementary temporal representations for brain disease diagnosis.

an initial coefficient hypergraph. Specifically, for the n-th ROI in window t, the sparse reconstruction coefficient is obtained by:

$$\min_{\alpha_n^t} \frac{1}{2}\|x_n^t - \hat{X}_n^t \alpha_n^t\|_2^2 + \lambda\|\alpha_n^t\|_1. \tag{1}$$

where $x_n^t \in \mathbb{R}^L$ denotes the BOLD signal of the n-th ROI in window t, $\hat{X}_n^t \in \mathbb{R}^{L\times(N-1)}$ represents the signals of all remaining ROIs except ROI n, and $\alpha_n^t \in \mathbb{R}^{N-1}$ denotes the sparse reconstruction coefficient vector describing the contribution of other ROIs to the target ROI. The obtained sparse reconstruction coefficients are converted into the node–hyperedge incidence matrix $H_t^{(0)}$, where each hyperedge represents the high-order functional relationship centered on one ROI.

High-order Encoding. Based on the initialized hypergraph, we employ hypergraph convolution to propagate high-order functional information among brain regions.The initial node feature matrix is defined as $Z_0^t = (X_t)^T$. The l-th hypergraph convolution layer is formulated as:

$$Z_l^t = \sigma\left(D_{v,t}^{-\frac{1}{2}} H_t^{(0)} D_{e,t}^{-1} (H_t^{(0)})^T D_{v,t}^{-\frac{1}{2}} Z_{l-1}^t \Theta_l + b_l\right), \tag{2}$$

where $D_{v,t}$ and $D_{e,t}$ denote the node-degree and hyperedge-degree matrices, respectively. After L layers of hypergraph convolution, the output Z_L^t is denoted as $\widetilde{Z}_t$, which serves as the initial high-order representation for subsequent topology evolution.

3.2 History-aware Topology Evolution

Topology Evidence Modeling. To model temporally continuous topology changes, we design a history-aware topology evidence mechanism that integrates neighboring-state transitions and multi-window historical dependencies to provide evidence for topology evolution. For the t-th temporal window, let $\widetilde{Z}_t$ denote the initial high-order representation from the current hypergraph, and Z_{t-1} denote the evolved representation of the previous window. For the first temporal window, Z_{t-1} is initialized as zero. We characterize neighboring-window state transitions by $\Delta Z_t = \widetilde{Z}_t - Z_{t-1}$ and generate a gating coefficient:

$$G_t = \sigma(\Delta Z_t W_g + b_g), \tag{3}$$

where $\sigma(\cdot)$ denotes the sigmoid function. The locally history-enhanced representation is obtained through residual aggregation:

$$\overline{Z}_t = \widetilde{Z}_t + G_t \odot Z_{t-1}. \tag{4}$$

To capture longer-term temporal dependencies, we further employ multi-window historical attention to select relevant information from preceding temporal states. Specifically, Using the node representation $\overline{z}_{t,i}$ of ROI i as query information, the historical attention weight between the current window t and a previous window j is calculated as:

$$\alpha_{t,i,j} = \frac{\exp(q_{t,i}^T k_{j,i}/\sqrt{d_h})}{\sum_{h\in\mathcal{H}_t}\exp(q_{t,i}^T k_{h,i}/\sqrt{d_h})}, \tag{5}$$

where i indexes ROI nodes, j denotes historical windows, and d_h represents the hidden dimension. The attention weights are computed independently for each ROI to select relevant historical states, whose weighted representations are aggregated as the historical context M_t and integrated with the current representation to generate topology evidence:

$$Z_t^{\text{top}} = \text{LayerNorm}\left(\overline{Z}_t + \gamma M_t\right), \tag{6}$$

where γ controls the contribution of historical information. The resulting Z_t^{top} serves as topology evidence for subsequent membership evolution.

Discrete Topology Evolution. Based on the topology evidence Z_t^{top}, we perform budget-constrained discrete topology evolution to adaptively refine node–hyperedge memberships. Specifically, we measure the functional consistency between candidate nodes and hyperedge centers in the topology evidence space. Let h_i^t and h_e^t denote the topology evidence embeddings of candidate node i and hyperedge center node e, respectively. The consistency score is computed as:

$$s_{i,e}^t = \frac{(h_i^t)^T h_e^t}{\|h_i^t\|_2 \|h_e^t\|_2}. \tag{7}$$

A high consistency score indicates a potential new functional association, whereas a low score suggests that an existing membership may become inconsistent with the evolving topology. Therefore, the editing scores for addition and deletion are defined as:

$$r_{i,e}^{add} = \frac{s_{i,e}^t}{\tau}, \qquad r_{i,e}^{del} = -\frac{s_{i,e}^t}{\tau}, \tag{8}$$

where τ controls the smoothness of editing scores. Under predefined topology budgets, Top-k selection is applied to the addition and deletion score matrices to obtain discrete editing masks:

$$D_t^{hard} = TopK(R_t^{del}), \qquad A_t^{hard} = TopK(R_t^{add}), \tag{9}$$

where R_t^{del} and R_t^{add} denote the deletion and addition score matrices, respectively. Since hard Top-k selection is non-differentiable, we employ a straight-through estimator (STE)Yin et al. (2019) to obtain differentiable topology editing masks:

$$D_t^{STE} = D_t^{soft} + \text{StopGrad}(D_t^{hard} - D_t^{soft}), \tag{10}$$

where D_t^{soft} denotes the continuous relaxation of deletion decisions. The same operation is applied to obtain A_t^{STE} for addition decisions.The evolved hypergraph topology is obtained by updating the node–hyperedge incidence matrix:

$$H_t' = H_t^{(0)} \odot (1 - D_t^{STE}) + (1 - H_t^{(0)}) \odot A_t^{STE}. \tag{11}$$

The first term preserves existing memberships after deletion, while the second term introduces new node–hyperedge connections. After topology evolution, the current-window features are re-encoded on the evolved hypergraph as $Z_t = \text{HGEnc}(X_t, H_t')$.

3.3 Kernel Attention Fusion

After topology evolution, the model obtains the evolved node representations $\{Z_t\}_{t=1}^W$ from different temporal windows. To integrate complementary information across temporal windows, we employ a Kernel Attention Fusion (KAF) Vaswani et al. (2017) module that adaptively aggregates

window-level representations while preserving high-order functional information. For each temporal window, the node representation is flattened and projected into a unified semantic space as $u_t = \mathrm{Dropout}[\mathrm{SiLU}(\mathrm{LN}(W_g\mathrm{Vec}(Z_t) + b_g))]$, where $\mathrm{Vec}(\cdot)$ denotes the vectorization operation. Based on the window representations, a global context query and window-specific keys are generated to measure the importance of each temporal window. The attention weight is computed as:

$$\beta_t = \frac{\exp(q^T k_t)}{\sum_{j=1}^{W} \exp(q^T k_j)}, \tag{12}$$

where q and k_t denote the global query and the key representation of window t, respectively. The fused temporal representation is obtained by aggregating window-level features:

$$u_{\mathrm{fus}} = \sum_{t=1}^{W} \beta_t u_t. \tag{13}$$

To preserve complementary high-order information across different temporal windows, the fused temporal representation is combined with node-level representations from all windows to form the final classification feature:

$$F_{\mathrm{cls}} = \mathrm{Concat}(\mathrm{Vec}(Z_1), \ldots, \mathrm{Vec}(Z_W), u_{\mathrm{fus}}). \tag{14}$$

where $\mathrm{Vec}(\cdot)$ denotes the vectorization operation and $\mathrm{Concat}(\cdot)$ denotes feature concatenation. Finally, the fused representation is fed into a fully connected classifier with Softmax activation for brain disease classification. The whole framework is optimized in an end-to-end manner using the standard cross-entropy loss.

4 EXPERIMENTS

Datasets. We evaluate HyperEvo on two public rs-fMRI datasets: the Alzheimer's Disease Neuroimaging Initiative (ADNI, `https://adni.loni.usc.edu/`) and the Parkinson's Progression Markers Initiative (PPMI, `https://www.ppmi-info.org/`). The selected ADNI data contain 245 samples, including 100 cognitively normal controls (CN), 84 patients with early mild cognitive impairment (EMCI), and 61 participants with subjective memory complaints (SMC). The selected PPMI data contain 156 samples, including 54 patients with Parkinson's disease (PD), 41 healthy controls (HC), and 61 participants from the prodromal cohort (Prodromal).

Implementation Details and Evaluation Metrics. All experiments are conducted using PyTorch 2.1.0 on an NVIDIA RTX 3090 GPU. We optimize the model with Adam using a batch size of 4, a learning rate of 5×10^{-4}, and the negative log-likelihood loss. Training runs for at most 300 epochs with early stopping (patience 50). We employ stratified 10-fold cross-validation Li et al. (2021) with identical partitions and random seeds for all methods. Performance is evaluated using ACC, AUC, and F1; macro one-vs-rest AUC and macro-F1 are used for multi-class tasks. Preprocessing, window construction, model settings, and task-specific hyperparameters are detailed in Appendix A.1.

Comparison Methods. To evaluate the effectiveness of HyperEvo, we compare it with representative conventional machine-learning and deep-learning methods for brain network analysis. The conventional methods include support vector machine (SVM) Cortes & Vapnik (1995), canonical correlation analysis (CCA) Hotelling (1992), and least absolute shrinkage and selection operator (LASSO) Jie et al. (2016). The deep-learning baselines include recurrent neural network (RNN) Cho et al. (2014), graph attention network (GAT) Veličković et al. (2017), long short-term memory network (LSTM) Graves (2012), Brain Modularity-constrained dynamic Representation learning (BMR) Wang et al. (2024), Neuro-Heterogeneity Guided Temporal Learning (NH-TGL) Li et al. (2026b), Asynchronous Functional Brain Network Construction (ACIFBN) Zhang et al. (2024), and Spatio-Temporal Hypergraph Attention Network (STHAN) Li et al. (2026a). The comparison results are presented in Table 1. For readability, Table 1 reports the mean ACC, AUC, and F1 scores over the ten cross-validation folds without standard deviations. The complete results with mean $\pm$ standard deviation are provided in Appendix A.2.

Ablation Experiment. To evaluate the contribution of each component in HyperEvo to classification performance, we conduct ablation experiments on multiple classification tasks from the ADNI and PPMI datasets. For each task, the complete model is used as the baseline, and only one component is removed or replaced at a time while keeping other experimental settings unchanged. The ablation results are presented in Table 2.

Table 1: Classification performance (%) on ADNI and PPMI datasets. Mean and standard deviation results are provided in the appendix.

Dataset	Method	NC vs. EMCI			NC vs. SMC			SMC vs. EMCI			NC vs. SMC vs. EMCI		
		ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
ADNI	SVM	67.37	75.24	64.77	66.43	72.38	59.67	69.71	81.67	73.57	56.32	74.35	55.80
	CCA	70.64	81.01	63.98	73.90	79.05	62.85	70.33	81.65	74.37	58.00	75.64	57.29
	LASSO	72.78	76.78	68.88	74.52	78.05	67.21	72.29	80.96	73.45	61.20	75.85	60.17
	RNN	70.64	68.53	67.86	80.11	74.02	70.58	80.95	85.94	84.59	52.65	63.57	49.08
	LSTM	70.06	69.17	64.76	81.36	73.71	69.11	79.48	86.78	82.96	56.80	68.92	53.47
	GAT	77.69	72.71	74.65	79.52	76.67	66.29	81.43	84.02	83.72	66.12	72.70	64.40
	BMR	83.15	76.74	78.60	81.91	74.47	67.65	82.04	77.82	83.71	68.08	74.13	66.39
	NH-TGL	83.59	78.96	80.62	80.56	74.08	71.30	81.07	76.20	84.11	63.30	70.71	60.99
	ACIFBN	84.27	82.40	80.69	78.86	76.69	67.15	81.38	80.27	81.64	64.79	71.70	60.03
	STHAN	85.35	87.40	82.69	84.56	78.71	74.84	84.19	84.79	86.59	68.50	74.63	64.30
	Ours	**88.16**	**89.67**	**85.65**	**85.22**	**85.79**	**79.77**	**89.67**	**90.12**	**90.95**	**73.55**	**81.82**	**72.06**
Dataset	**Method**	**PD vs. HC**			**PD vs. Prodromal**			**HC vs. Prodromal**			**PD vs. HC vs. Prodromal**		
		ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
PPMI	SVM	59.44	60.32	45.94	57.42	61.02	50.50	59.73	55.19	70.07	46.50	63.43	41.74
	CCA	62.67	60.02	46.94	64.14	65.14	61.72	58.73	57.48	66.43	49.58	65.41	48.64
	LASSO	61.78	54.09	50.37	64.24	69.33	63.22	62.73	71.80	62.80	46.83	61.33	41.82
	RNN	66.33	62.93	57.42	70.61	64.85	73.41	63.73	55.73	66.28	47.92	56.93	42.98
	LSTM	68.44	61.27	59.32	62.58	53.15	61.28	65.73	61.76	69.95	48.79	57.29	44.91
	GAT	70.67	65.50	49.84	69.62	64.11	71.62	76.71	64.31	80.01	52.04	59.02	42.88
	BMR	73.78	63.38	56.67	70.61	60.60	68.60	76.64	67.35	81.34	55.92	61.74	49.12
	NH-TGL	76.67	72.80	66.89	73.86	70.11	76.16	77.55	70.32	82.36	57.79	64.88	54.05
	ACIFBN	76.67	61.45	67.31	72.28	67.89	74.64	75.55	68.51	79.82	55.21	58.52	46.72
	STHAN	81.33	74.17	74.89	81.89	75.84	82.97	78.55	68.73	**85.00**	60.42	62.40	55.80
	Ours	**85.22**	**80.17**	**88.41**	**83.56**	**79.49**	**84.67**	**79.64**	**75.46**	83.39	**62.17**	**66.92**	**60.48**

- **History-aware topology evidence (HTE).** Removing HTE reduces ACC by 2.10–9.37 percentage points (p.p.) on ADNI and 4.54–5.33 p.p. on PPMI, demonstrating the value of historical states for current topology evolution. The separate removal of the adjacent-window gate ($\mathrm{HTE}_{\mathrm{gate}}$) or history attention ($\mathrm{HTE}_{\mathrm{attn}}$) also causes task-dependent degradation, indicating that local temporal continuity and longer-range history provide complementary topology evidence.
- **Discrete topology evolution (DTE).** Disabling topology evolution decreases ACC by 0.85–4.10 p.p. on ADNI and 4.00–14.77 p.p. on PPMI. Removing the straight-through estimator (STE) further causes reductions of 1.25–4.41 and 3.19–6.14 p.p., respectively. These results show that both explicit membership editing and its surrogate gradient are important for learning task-relevant topology changes.
- **Window-level kernel attention fusion (WKAF).** Replacing WKAF with direct flattening decreases ACC in every task, with drops of 6.61–9.62 p.p. on ADNI and 4.01–10.91 p.p. on PPMI. This confirms that selectively aggregating discriminative windows is more effective than treating all temporal states equally. Although a few ablations produce isolated improvements in AUC or F1 on PPMI, the complete model achieves the highest ACC consistently across all eight tasks.

5 Discussion

The parameter sensitivity analyses and brain network visualizations on ADNI are presented in the main text, while the corresponding results on PPMI are provided in the Appendix A.3.

Window Sensitivity Analysis. We investigate the effects of the number of temporal windows $W \in \{3-10\}$ and the history length $K \in \{0-4\}$. As shown in Fig. 3, the optimal number of windows varies across tasks: $W = 5$ for CN vs. EMCI, $W = 8$ for CN vs. SMC, and $W = 7$ for both SMC vs. EMCI and the three-class task. Too few windows may obscure short-term dynamics, whereas

Table 2: Ablation results of HyperEvo on ADNI and PPMI classification tasks (%)

Dataset	Method	NC vs. EMCI			NC vs. SMC			SMC vs. EMCI			NC vs. SMC vs. EMCI		
		ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
ADNI	w/o HTE	81.61	85.81	78.17	75.85	73.76	67.37	87.57	88.40	89.57	71.10	78.85	69.18
	w/o HTE_{gate}	85.96	87.54	83.84	80.81	76.14	73.73	87.57	87.41	89.01	73.15	80.48	71.28
	w/o HTE_{attn}	83.22	84.81	79.86	84.52	85.02	78.84	83.48	85.04	86.11	71.83	79.45	69.75
	w/o DTE	84.30	85.39	80.19	81.47	78.81	72.24	85.57	87.27	87.64	72.70	81.17	72.03
	w/o STE	85.41	88.92	83.08	80.81	77.57	71.41	87.57	88.40	89.57	72.30	77.42	70.07
	w/o WKAF	81.55	82.51	77.54	77.76	71.17	67.93	80.05	80.93	83.93	66.15	77.95	64.25
	Ours (complete)	**88.16**	**89.67**	**85.65**	**85.22**	**85.79**	**79.77**	**89.67**	**90.12**	**90.95**	**73.55**	**81.82**	**72.06**
Dataset	Method	PD vs. HC			PD vs. Prodromal			HC vs. Prodromal			PD vs. HC vs. Prodromal		
		ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
PPMI	w/o HTE	79.89	77.89	84.68	78.48	72.51	77.77	74.73	72.38	79.18	57.67	63.36	54.33
	w/o HTE_{gate}	77.89	**82.50**	82.16	78.41	72.25	78.19	77.36	75.43	81.71	58.29	**67.32**	55.53
	w/o HTE_{attn}	81.22	75.00	84.31	82.65	**81.73**	84.27	76.64	71.75	80.91	55.00	59.45	51.93
	w/o DTE	77.89	73.08	82.35	68.79	68.32	68.61	75.64	71.95	78.75	53.87	63.55	52.05
	w/o STE	79.67	79.52	83.68	77.42	67.48	79.24	76.45	73.17	81.33	57.58	62.58	54.40
	w/o WLAF	80.11	73.95	83.56	79.55	75.59	81.21	68.73	66.67	75.42	57.54	63.51	**65.04**
	Ours (complete)	**85.22**	80.17	**88.41**	**83.56**	79.49	**84.67**	**79.64**	**75.46**	**83.39**	**62.21**	67.00	**58.12**

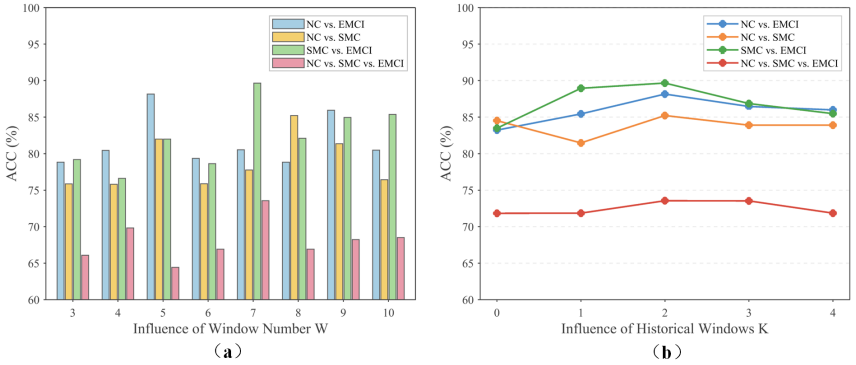

Figure 3: Window sensitivity analysis on ADNI. (a) Effect of the number of temporal windows ($W \in \{3-10\}$). (b) Effect of the history length ($K \in \{0-4\}$).

excessive partitioning reduces the number of time points per window and may destabilize high-order relation estimation. In contrast, all four tasks achieve their best performance at $K = 2$, indicating that a moderate history length provides sufficient temporal context without introducing redundant information.

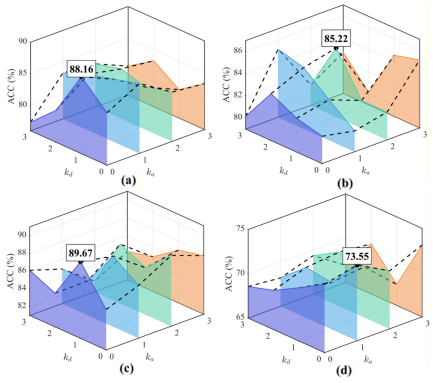

Figure 4: Effect of hypergraph topology editing scores on classification performance. (a) NC vs. EMCI. (b) NC vs. SMC. (c) SMC vs. EMCI. (d) NC vs. SMC vs. EMCI.

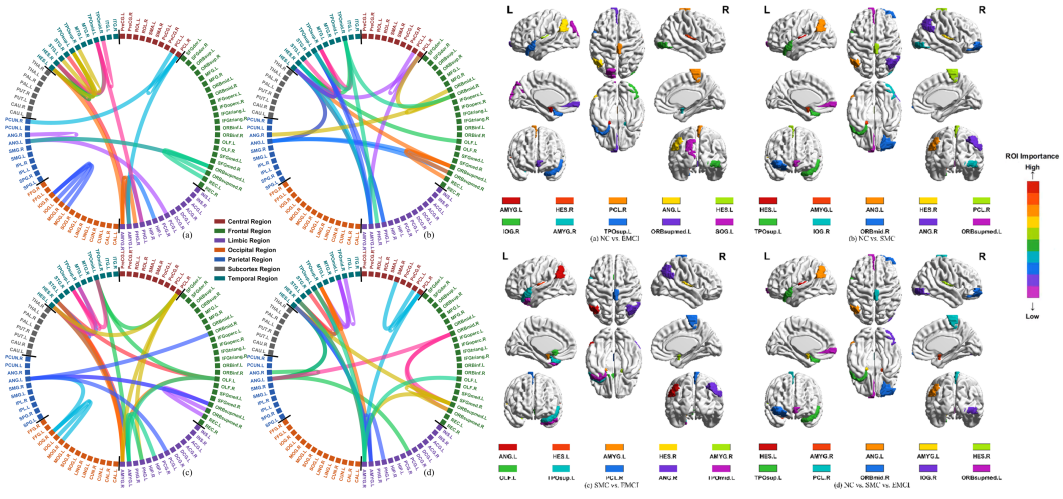

Figure 5: Brain network interpretability on ADNI.The left and right sides of each panel show the top-10 discriminativehyperedges and ROIs, respectively:(a) NC vs. EMCI, (b) NC vs. SMC,(c) SMC vs. EMCI, and (d) NC vs. SMC vs. EMCI.

Topology Editing Sensitivity Analysis. We vary the editing scores for addition and deletion, $k_a, k_d \in \{0-3\}$, to examine the effect of the intensity of the topology evolution. As shown in Fig. 4, the optimal (k_a, k_d) differs across tasks: $(0, 1)$ for NC vs. EMCI and SMC vs. EMCI, $(2, 2)$ for NC vs. SMC, and $(1, 0)$ for the three-class task. These results indicate that task-specific moderate editing improves classification, whereas excessive structural changes may disrupt stable high-order relationships in the initial hypergraph.

Brain Network Interpretability. Figure 5 presents the top-10 discriminative hyperedges (left) and ROIs ranked by class-wise membership differences (right). In Fig. 5(a), the AMYG–HIP–PHG hyperedge and the highly ranked bilateral Heschl's gyri and amygdalae indicate limbic–memory and auditory–limbic alterationsin EMCI (Yao et al., 2014; Li et al., 2016; Wang et al., 2020). In Fig. 5(b), the HES–TPO–AMYG–STG hyperedge and the prominent Heschl's, angular, temporal-pole, and orbitofrontal regions suggest auditory–temporal, limbic, and frontoparietal reorganization in SMC (Sun et al., 2016; Huang et al., 2024; Paban et al., 2023). In Fig. 5(c), the recurring HES–AMYG–TPO hyperedges, together with olfactory, angular, and paracentral regions, suggest distributed sensory–limbic–associative differences between SMC and EMCI. This finding is broadly consistent with the convergent and divergent connectivity alterations reported between the clinically related SCD and aMCI populations (Xue et al., 2021; Cai et al., 2020; López-Sanz et al., 2017). In Fig. 5(d), the recurrent HES–AMYG–TPO pattern and the additional involvement of angular and orbitofrontal regions indicate a more distributed topology for three-class discrimination. These results broadly align with multiscale network alterations and distinct subnetwork connectivity reported across the preclinical AD spectrum (Wang et al., 2019; Yuan et al., 2022; López-Sanz et al., 2017).

6 Conclusion

We proposed HyperEvo, a history-aware framework that learns dynamic hypergraphs for brain disease classification. It combines historical evidence, budget-constrained discrete topology editing, and window-level attention to model evolving high-order brain networks. Experiments and ablations across eight classification tasks on ADNI and PPMI demonstrate its effectiveness. Interpretability analyses reveal task-specific brain regions, connections, and hyperedges associated with disease-related topological reorganization. Future work will explore adaptive temporal scales and evaluate generalization to multi-center and multimodal data.

AI USE STATEMENT

In this work, we used large language model-based generative AI tools to assist with research ideation, experimental configuration, software development and debugging, literature discovery, drafting selected passages, and language refinement. We did not use generative AI to generate synthetic datasets or establish mathematical proofs, and these uses are not applicable to this work. All AI-assisted outputs were critically reviewed by the authors. Suggested references and scientific claims were verified against the original literature, AI-assisted code was manually inspected and tested, and experimental configurations, results, and interpretations were independently checked by the authors. Generative AI was not used as a source of scientific evidence or to make final research decisions. The authors take full responsibility for the final content of this work, including all text, claims, code, and artifacts produced with the assistance of generative AI.

ETHICS STATEMENT

This study involves secondary analysis of de-identified rs-fMRI data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Parkinson's Progression Markers Initiative (PPMI). Data were accessed and used in accordance with the respective data-use policies. The original studies obtained approval from their participating institutions and informed consent from all participants or their authorized representatives. This work involved no new participant recruitment, data collection, or clinical intervention. We report group-level results only and do not attempt to identify individual participants.

REFERENCES

Chunting Cai, Chenxi Huang, Chenhui Yang, Haijie Lu, Xin Hong, Fujia Ren, Dan Hong, and Eyk Ng. Altered patterns of functional connectivity and causal connectivity in salience subnetwork of subjective cognitive decline and amnestic mild cognitive impairment. *Frontiers in Neuroscience*, 14:288, 2020.

Claire HC Chang, Samuel A Nastase, and Uri Hasson. Information flow across the cortical timescale hierarchy during narrative construction. *Proceedings of the National Academy of Sciences*, 119 (51):e2209307119, 2022.

Kyunghyun Cho, Bart Van Merriënboer, Çağlar Gulçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, 2014.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

Eran Dayan and Nina Browner. Alterations in striato-thalamo-pallidal intrinsic functional connectivity as a prodrome of Parkinson's disease. *NeuroImage: Clinical*, 16:313–318, 2017.

Harini Eavani, Theodore D Satterthwaite, Roman Filipovych, Raquel E Gur, Ruben C Gur, and Christos Davatzikos. Identifying sparse connectivity patterns in the brain using resting-state fmri. *NeuroImage*, 105:286–299, 2015.

Charalampos Georgiopoulos, Suzanne T Witt, Sven Haller, Nil Dizdar, Helene Zachrisson, Maria Engström, and Elna-Marie Larsson. A study of neural activity and functional connectivity within the olfactory brain network in Parkinson's disease. *NeuroImage: Clinical*, 23:101946, 2019.

Alex Graves. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pp. 37–45, 2012.

Shi Gu, Muzhi Yang, John D Medaglia, Ruben C Gur, Raquel E Gur, Theodore D Satterthwaite, and Danielle S Bassett. Functional hypergraph uncovers novel covariant structures over neurodevelopment. *Human Brain Mapping*, 38(8):3823–3835, 2017.

Xiangmin Han, Rundong Xue, Jingxi Feng, Yifan Feng, Shaoyi Du, Jun Shi, and Yue Gao. Hypergraph foundation model for brain disease diagnosis. *IEEE Transactions on Neural Networks and Learning Systems*, 36(10):17702–17716, 2025.

Harold Hotelling. Relations between two sets of variates. In *Breakthroughs in Statistics: Methodology and Distribution*, pp. 162–190. Springer, 1992.

Wenqi Hu, Xuerui Su, Guanliang Li, Yidi Pan, and Aijing Lin. Learning high-order relationships with hypergraph attention-based spatio-temporal aggregation for brain disease analysis. *IEEE Journal of Biomedical and Health Informatics*, 2026.

Shaochun Huang, Siyu Wang, Zigang Che, Honglin Ge, Zheng Yan, Jia Fan, Xiang Lu, Li Liu, Wan Liu, Yeming Zhong, et al. Brain-wide functional connectivity alterations and their cognitive correlates in subjective cognitive decline. *Frontiers in Neuroscience*, 18:1438260, 2024.

Junzhong Ji and Yaqin Zhang. Functional brain network classification based on deep graph hashing learning. *IEEE Transactions on Medical Imaging*, 41(10):2891–2902, 2022.

Biao Jie, Chong-Yaw Wee, Dinggang Shen, and Daoqiang Zhang. Hyper-connectivity of functional networks for brain disease diagnosis. *Medical Image Analysis*, 32:84–100, 2016.

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

Chaojun Li, Peiliang Gong, Shengrong Li, Chunwei Tian, Yinbo Yu, Ran Wang, Daoqiang Zhang, and Qi Zhu. Spatio-temporal hypergraph attention networks for brain disease analysis. *IEEE Transactions on Image Processing*, 2026a.

Shengrong Li, Qi Zhu, Donghai Guan, Bo Shen, Li Zhang, Yixin Ji, Shile Qi, and Daoqiang Zhang. Long-interval spatio-temporal graph convolution for brain disease diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 74:1–11, 2025a.

Shengrong Li, Qi Zhu, Chunwei Tian, Li Zhang, Bo Shen, Chuhang Zheng, Daoqiang Zhang, and Wei Shao. Spatio-temporal evolutionary graph learning for brain network analysis using medical imaging. *IEEE Transactions on Image Processing*, 2025b.

Shengrong Li, Qi Zhu, Chunwei Tian, Xinyang Zhang, Wei Shao, Jie Wen, and Daoqiang Zhang. Neuroh-tgl: Neuro-heterogeneity guided temporal graph learning strategy for brain disease diagnosis. *Advances in Neural Information Processing Systems*, 38:109349–109371, 2026b.

Yang Li, Jingyu Liu, Yiqiao Jiang, Yu Liu, and Baiying Lei. Virtual adversarial training-based deep feature aggregation network from dynamic effective connectivity for mci identification. *IEEE Transactions on Medical Imaging*, 41(1):237–251, 2021.

Yuxia Li, Xiaoni Wang, Yongqiu Li, Yu Sun, Can Sheng, Hongyan Li, Xuanyu Li, Yang Yu, Guanqun Chen, Xiaochen Hu, et al. Abnormal resting-state functional connectivity strength in mild cognitive impairment and its conversion to Alzheimer's disease. *Neural Plasticity*, 2016(1): 4680972, 2016.

Lingwen Liu, Guangqi Wen, Peng Cao, Tianshun Hong, Jinzhu Yang, Xizhe Zhang, and Osmar R Zaiane. Braintgl: A dynamic graph representation learning model for brain network analysis. *Computers in Biology and Medicine*, 153:106521, 2023.

Shu-Ying Liu, Piu Chan, and A. Jon Stoessl. The underlying mechanism of prodromal pd: insights from the parasympathetic nervous system and the olfactory system. *Translational Neurodegeneration*, 6:4, 2017.

David López-Sanz, Ricardo Bruña, Pilar Garcés, María Carmen Martín-Buro, Stefan Walter, María Luisa Delgado, Mercedes Montenegro, Ramon Lopez Higes, Alberto Marcos, and Fernando Maestú. Functional connectivity disruption in subjective cognitive decline and mild cognitive impairment: a common pattern of alterations. *Frontiers in Aging Neuroscience*, 9:109, 2017.

Yulan Ma, Fangkun Li, Wenchao Yang, Qian Si, Chenglong Yu, and Yang Li. Hyperdiag: Temporal–regional hypergraph learning via topology-enhanced state propagation for brain disease diagnosis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 7874–7882, 2026.

Véronique Paban, Ahmad Mheich, Laure Spieser, and Mathilde Sacher. A multidimensional model of memory complaints in older individuals and the associated hub regions. *Frontiers in Aging Neuroscience*, 15:1324309, 2023.

Xiaoying Song, Ke Wu, and Li Chai. Brain network analysis of schizophrenia patients based on hypergraph signal processing. *IEEE Transactions on Image Processing*, 32:4964–4976, 2023.

Yu Sun, Zhengjia Dai, Yuxia Li, Can Sheng, Hongyan Li, Xiaoni Wang, Xiaodan Chen, Yong He, and Ying Han. Subjective cognitive decline: mapping functional and structural brain changes—a combined resting-state functional and structural MR imaging study. *Radiology*, 281(1):185–192, 2016.

Konrad Szewczyk-Krolikowski, Ricarda AL Menke, Michal Rolinski, Eugene Duff, Gholamreza Salimi-Khorshidi, Nicola Filippini, Giovanna Zamboni, Michele TM Hu, and Clare E Mackay. Functional connectivity in the basal ganglia network differentiates pd patients from controls. *Neurology*, 83(3):208–214, 2014.

Yingzhi Teng, Kai Wu, Jing Liu, Yifan Li, and Xiangyi Teng. Constructing high-order functional connectivity networks with temporal information from fmri data. *IEEE Transactions on Medical Imaging*, 43(12):4133–4145, 2024.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

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

Diana Wang, Alexander Belden, Suzanne B Hanser, Maiya R Geddes, and Psyche Loui. Resting-state connectivity of auditory and reward systems in Alzheimer's disease and mild cognitive impairment. *Frontiers in Human Neuroscience*, 14:280, 2020.

Junling Wang, Lianglong Sun, Lili Chen, Junyan Sun, Yapei Xie, Dezheng Tian, Linlin Gao, Dongling Zhang, Mingrui Xia, and Tao Wu. Common and distinct roles of amygdala subregional functional connectivity in non-motor symptoms of Parkinson's disease. *npj Parkinson's Disease*, 9(1):28, 2023a.

Junqi Wang, Hailong Li, Gang Qu, Kim M Cecil, Jonathan R Dillman, Nehal A Parikh, and Lili He. Dynamic weighted hypergraph convolutional network for brain functional connectome analysis. *Medical Image Analysis*, 87:102828, 2023b.

Mingliang Wang, Jiashuang Huang, Mingxia Liu, and Daoqiang Zhang. Modeling dynamic characteristics of brain functional connectivity networks using resting-state functional mri. *Medical Image Analysis*, 71:102063, 2021.

Qianqian Wang, Wei Wang, Yuqi Fang, Pew-Thian Yap, Hongtu Zhu, Hong-Jun Li, Lishan Qiao, and Mingxia Liu. Leveraging brain modularity prior for interpretable representation learning of fmri. *IEEE Transactions on Biomedical Engineering*, 71(8):2391–2401, 2024.

Ziqi Wang, Kaini Qiao, Guanqun Chen, Danyang Sui, Hao-Ming Dong, Yin-Shan Wang, Hui-Jie Li, Jie Lu, Xi-Nian Zuo, and Ying Han. Functional connectivity changes across the spectrum of subjective cognitive decline, amnestic mild cognitive impairment and Alzheimer's disease. *Frontiers in Neuroinformatics*, 13:26, 2019.

Chong-Yaw Wee, Pew-Thian Yap, Daoqiang Zhang, Lihong Wang, and Dinggang Shen. Group-constrained sparse fmri connectivity modeling for mild cognitive impairment identification. *Brain Structure and Function*, 219(2):641–656, 2014.

Guangqi Wen, Peng Cao, Lingwen Liu, Maochun Hao, Siyu Liu, Junjie Zheng, Jinzhu Yang, Osmar R Zaiane, and Fei Wang. Heterogeneous graph representation learning framework for resting-state functional connectivity analysis. *IEEE Transactions on Medical Imaging*, 44(3):1581–1595, 2024.

Chen Xue, Wei Qi, and Qing Yuan. Disrupted dynamic functional connectivity in distinguishing subjective cognitive decline and amnestic mild cognitive impairment based on the triple-network model. *Frontiers in Aging Neuroscience*, 13:711009, 2021.

Hongxiang Yao, Bo Zhou, Zengqiang Zhang, Pan Wang, Yan'e Guo, Yanchang Shang, Luning Wang, Xi Zhang, Ningyu An, Yong Liu, et al. Longitudinal alteration of amygdalar functional connectivity in mild cognitive impairment subjects revealed by resting-state fMRI. *Brain Connectivity*, 4(5):361–370, 2014.

Penghang Yin, Jiancheng Lyu, Shuai Zhang, Stanley Osher, Yingyong Qi, and Jack Xin. Understanding straight-through estimator in training activation quantized neural nets, 2019.

Qianqian Yuan, Xuhong Liang, Chen Xue, Wenzhang Qi, Shanshan Chen, Yu Song, Huimin Wu, Xulian Zhang, Chaoyong Xiao, and Jiu Chen. Altered anterior cingulate cortex subregional connectivity associated with cognitions for distinguishing the spectrum of pre-clinical Alzheimer's disease. *Frontiers in Aging Neuroscience*, 14:1035746, 2022.

Weiqi Zeng, Wenliang Fan, Xiangchuang Kong, Xiaoming Liu, Ling Liu, Ziqin Cao, Xiaoqian Zhang, Xiaoman Yang, Chi Cheng, Yi Wu, et al. Altered intra- and inter-network connectivity in drug-naïve patients with early parkinson's disease. *Frontiers in Aging Neuroscience*, 14:783634, 2022.

Jianjia Zhang, Xiaotong Wu, Xiang Tang, Luping Zhou, Lei Wang, Weiwen Wu, and Dinggang Shen. Asynchronous functional brain network construction with spatiotemporal transformer for mci classification. *IEEE Transactions on Medical Imaging*, 44(3):1168–1180, 2024.

Jie Zhou, Biao Jie, Zhengdong Wang, Zhixiang Zhang, Weixin Bian, Yang Yang, Hongwei Li, Fengyun Sun, and Mingxia Liu. Dclnet: Double collaborative learning network on stationary-dynamic functional brain network for brain disease classification. *IEEE Transactions on Image Processing*, 34:4026–4039, 2025.

Qi Zhu, Ruting Xu, Ran Wang, Xijia Xu, Zhiqiang Zhang, and Daoqiang Zhang. Stacked topological preserving dynamic brain networks representation and classification. *IEEE Transactions on Medical Imaging*, 41(11):3473–3484, 2022.

Qi Zhu, Shengrong Li, Xiangshui Meng, Qiang Xu, Zhiqiang Zhang, Wei Shao, and Daoqiang Zhang. Spatio-temporal graph hubness propagation model for dynamic brain network classification. *IEEE Transactions on Medical Imaging*, 43(6):2381–2394, 2024.

Chen Zu, Yue Gao, Brent Munsell, Minjeong Kim, Ziwen Peng, Yingying Zhu, Wei Gao, Daoqiang Zhang, Dinggang Shen, and Guorong Wu. Identifying high order brain connectome biomarkers via learning on hypergraph. In *International Workshop on Machine Learning in Medical Imaging*, pp. 1–9, 2016.

A APPENDIX

A.1 TASK-SPECIFIC SPARSE REGULARIZATION

Preprocessing and Hypergraph Construction. For each scan, every ROI time series is standardized independently and divided into continuous, equal-length, non-overlapping windows. Trailing time points that cannot form a complete window are discarded. A sparse hypergraph is then independently constructed for each window using LASSO regression.

Model and Training Settings. The hidden dimension is set to 8, with two hypergraph convolution layers before and after topology evolution. We employ the WindowToken32 linear-kernel attention classifier. Models are optimized using Adam for at most 300 epochs with a batch size of 4, a learning rate of 5×10^{-4}, and early stopping with a patience of 50. The base random seed is fixed to 7, and identical cross-validation partitions are used for all methods.

Task-Specific Sparsity Settings. The LASSO regularization coefficient λ controls the sparsity of the initial window-level hypergraphs. Table 3 reports the coefficient used for each classification task.

Table 3: Task-specific LASSO regularization coefficients for initial sparse hypergraph construction.

DATASET	CLASSIFICATION TASK	λ
ADNI	CN vs. EMCI	0.3
	CN vs. SMC	0.1
	SMC vs. EMCI	0.1
	CN vs. SMC vs. EMCI	0.1
PPMI	PD vs. HC	0.4
	PD vs. Prodromal	0.4
	HC vs. Prodromal	0.2
	PD vs. HC vs. Prodromal	0.1

A.2 COMPLETE CLASSIFICATION RESULTS

Tables 4–7report the complete classification results on ADNI and PPMI. All results are presented as mean/standard deviation (%) over the ten cross-validation folds, and the best mean value for each metric is shown in bold. Blank entries indicate unavailable results.

Table 4: Classification performance (%) on the NC vs. EMCI and NC vs. SMC tasks.

Method	NC vs. EMCI			NC vs. SMC		
	ACC	AUC	F1	ACC	AUC	F1
SVM	67.37/10.12	75.24/10.95	64.77/11.19	66.43/09.83	72.38/09.94	59.67/13.16
CCA	70.64/05.40	81.01/05.68	63.98/09.93	73.90/10.41	79.05/09.18	62.85/14.33
LASSO	72.78/08.56	76.78/11.50	68.88/14.39	74.52/10.27	78.05/11.19	67.21/11.76
RNN	70.64/08.14	68.53/08.04	67.86/06.58	80.11/08.77	74.02/15.41	70.58/14.84
LSTM	70.06/08.81	69.17/09.83	64.76/11.50	81.36/06.26	73.71/11.44	69.11/12.75
GAT	77.69/07.15	72.71/12.18	74.65/09.83	79.52/07.35	76.67/08.94	66.29/18.40
BMR	83.15/05.45	76.74/07.50	78.60/09.79	81.91/08.62	74.47/15.14	67.65/22.60
NH-TGL	83.59/05.95	78.96/09.27	80.62/07.18	80.56/06.32	74.08/15.03	71.30/12.45
ACIFBN	84.27/03.60	82.40/06.53	80.69/04.34	78.86/08.96	76.69/14.80	67.15/17.43
STHAN	85.35/04.37	87.40/07.96	82.69/05.60	84.56/06.59	78.71/11.43	74.84/13.07
Ours	**88.16/08.12**	**89.67/09.49**	**85.65/10.85**	**85.22/09.68**	**85.79/11.48**	**79.77/12.91**

A.3 ADDITIONAL RESULTS ON PPMI

To complement the ADNI analyses presented in the main text, this section provides additional results on the PPMI dataset. We examine the sensitivity of HyperEvo to the number of temporal

Table 5: Classification performance (%) on the SMC vs. EMCI and NC vs. SMC vs. EMCI tasks.

Method	SMC vs. EMCI			NC vs. SMC vs. EMCI		
	ACC	AUC	F1	ACC	AUC	F1
SVM	69.71/09.18	81.67/08.76	73.57/09.18	56.32/05.53	74.35/05.52	55.80/04.94
CCA	70.33/09.20	81.65/08.76	74.37/11.36	58.00/06.98	75.64/04.51	57.29/08.17
LASSO	72.29/11.17	80.96/09.92	73.45/14.20	61.20/06.66	75.85/08.05	60.17/08.05
RNN	80.95/13.22	85.94/12.04	84.59/09.83	52.65/06.55	63.57/06.54	49.08/07.10
LSTM	79.48/09.38	86.78/08.13	82.96/07.66	56.80/10.83	68.92/10.62	53.47/12.41
GAT	81.43/12.42	84.02/16.37	83.72/10.58	66.12/09.06	72.70/07.04	64.40/10.17
BMR	82.04/06.65	77.82/12.67	83.71/06.56	68.08/06.98	74.13/06.12	66.39/07.42
NH-TGL	81.07/04.62	76.20/09.73	84.11/04.74	63.30/07.00	70.71/06.45	60.99/07.82
ACIFBN	81.38/08.26	80.27/11.20	81.64/09.76	64.79/04.43	71.70/04.61	60.03/07.80
STHAN	84.19/08.06	84.79/07.43	86.59/06.61	68.50/07.62	74.63/08.02	64.30/08.52
Ours	**89.67/05.54**	**90.12/07.94**	**90.95/04.98**	**73.55/06.76**	**81.82/06.23**	**72.06/06.91**

Table 6: Classification performance (%) on the PD vs. HC and PD vs. Prodromal tasks.

Method	PD vs. HC			PD vs. Prodromal		
	ACC	AUC	F1	ACC	AUC	F1
SVM	59.44/14.60	60.32/20.53	45.94/24.79	57.42/08.62	61.02/09.26	50.50/13.27
CCA	62.67/14.35	60.02/29.18	46.94/24.59	64.14/09.81	65.14/13.75	61.72/15.13
LASSO	61.78/10.99	54.09/16.46	50.37/21.23	64.24/09.92	69.33/10.24	63.22/14.53
RNN	66.33/12.09	62.93/18.31	57.42/22.74	70.61/09.71	64.85/16.95	73.41/08.79
LSTM	68.44/11.29	61.27/19.86	59.32/24.34	62.58/11.37	53.15/15.92	61.28/23.37
GAT	70.67/09.97	65.50/14.57	49.84/28.78	69.62/06.45	64.11/11.76	71.62/06.69
BMR	73.78/08.55	63.38/19.93	56.67/26.42	70.61/07.45	60.60/17.12	68.60/12.08
NH-TGL	76.67/12.83	72.80/14.46	66.89/26.02	73.86/05.86	70.11/07.00	76.16/06.55
ACIFBN	76.67/10.52	61.45/15.08	67.31/24.51	72.28/06.05	67.89/11.76	74.64/05.40
STHAN	81.33/13.41	74.17/18.41	74.89/19.04	81.89/08.82	75.84/13.91	82.97/09.91
Ours	**85.22/09.60**	**80.17/13.32**	**88.41/07.18**	**83.56/08.64**	**79.49/14.52**	**84.67/07.76**

Table 7: Classification performance (%) on the HC vs. Prodromal and PD vs. HC vs. Prodromal tasks.

Method	HC vs. Prodromal			PD vs. HC vs. Prodromal		
	ACC	AUC	F1	ACC	AUC	F1
SVM	59.73/13.37	55.19/19.19	70.07/11.34	46.50/18.11	63.43/17.02	41.74/19.88
CCA	58.73/17.82	57.48/16.82	66.43/16.30	49.58/14.69	65.41/13.80	48.64/15.06
LASSO	62.73/16.48	71.80/15.56	62.80/20.10	46.83/08.05	61.33/10.11	41.82/08.66
RNN	63.73/14.10	55.73/16.90	66.28/16.98	47.92/12.08	56.93/10.59	42.98/13.23
LSTM	65.73/11.68	61.76/15.57	69.95/12.88	48.79/07.09	57.29/09.45	44.91/10.30
GAT	76.71/06.22	64.31/09.24	80.01/09.24	52.04/09.48	59.02/09.98	42.88/13.85
BMR	76.64/09.65	67.35/20.86	81.34/07.60	55.92/08.36	61.74/12.30	49.12/13.01
NH-TGL	77.55/07.39	70.32/12.22	82.36/05.98	57.79/09.72	64.88/09.30	54.05/11.79
ACIFBN	75.55/06.38	68.51/10.97	79.82/06.13	55.21/03.96	58.52/07.70	46.72/05.30
STHAN	78.55/08.06	68.73/15.51	**85.00/05.15**	60.42/07.57	62.40/09.01	55.80/12.06
Ours	**79.64/08.48**	**75.46/12.55**	83.39/07.52	**62.17/05.89**	**66.92/08.90**	**60.48/07.43**

windows, history length, and topology editing scores, followed by brain network interpretability analyses across the four PPMI classification tasks. All results in this section are reported as mean $\pm$ standard deviation across 10-fold cross-validation.

ppmi window sensitivity. We investigate the effect of the number of temporal windows $W \in \{3 - 10\}$ on the four PPMI classification tasks. As shown in Fig. 6, the optimal temporal scale varies across tasks. PD vs. HC and HC vs. Prodromal achieve their highest ACC values at $W = 5$, reaching 85.22% and 79.64%, respectively. PD vs. Prodromal performs best at $W = 8$ with an ACC of 83.56%, whereas the three-class task achieves its highest ACC of 62.21% at $W = 10$. These results indicate that different disease-state comparisons require different temporal resolutions. In particular, the three-class task benefits from finer temporal partitioning, which may help distinguish more heterogeneous functional states. Nevertheless, performance does not increase monotonically with W, since excessive partitioning also reduces the number of observations available for estimating each window-level hypergraph.

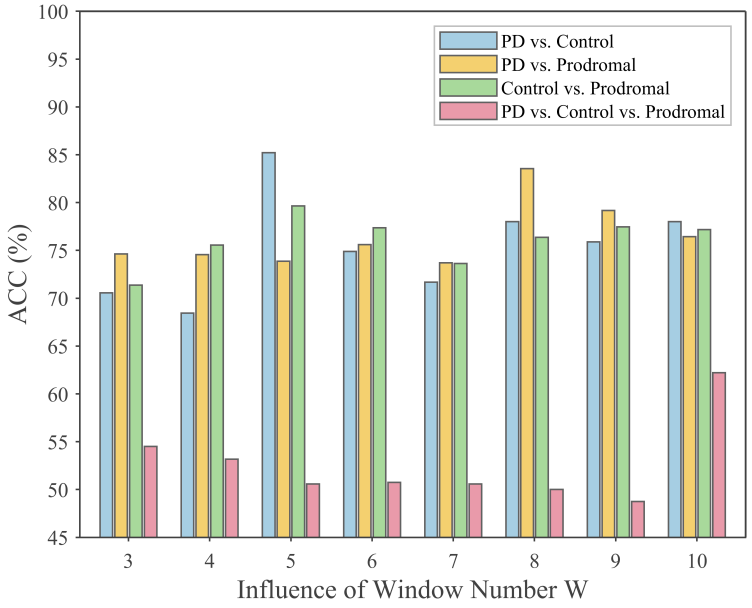

Figure 6: Effect of the number of temporal windows on PPMI.

History Length Sensitivity. We evaluate the history length $K \in \{0 - 4\}$ to investigate howmuch preceding temporal information is required for topology evolution. As shown in Fig. 7, all three binary tasks achieve their highest ACC at $K = 2$: 85.22% for PD vs. HC, 83.56% for PD vs. Prodromal, and 79.64% for HC vs. Prodromal. In contrast, the three-class task performs best at $K = 3$, reaching 62.17%. Compared with $K = 0$, incorporating historical information substantially improves performance, confirming its importance for topology evolution. However, increasing K beyond the optimum causes performance degradation, suggesting that excessively long histories may introduce redundant or less relevant temporal information.

Topology Editing Sensitivity. We vary the editing scores for addition and deletion, $k_a, k_d \in \{0-3\}$, to evaluate the effect of topology evolution intensity. As shown in Fig. 8, the optimal editing scores are task dependent. PD vs. HC achieves its highest ACC of 85.22% at $(k_a, k_d) = (3, 3)$, while PD vs. Prodromal reaches 83.56% at $(3, 2)$. Both HC vs. Prodromal and the three-class task perform best at $(2, 1)$, obtaining 79.64% and 62.21%, respectively. Compared with the fixed-topology setting $(0, 0)$, the optimal editing configuration improves ACC by 7.33, 14.77, 4.00, and 8.34 percentage points across the four tasks. These results demonstrate the benefit of topology evolution while indicating that its intensity should be adapted to the classification task.

Discriminative Hyperedges. Figure 9 visualizes the top-10 discriminative hyperedges for each PPMI task. In Fig. 9(a), the AMYG.L–PHG.L–ORBsupmed.L–AMYG.R–TPOsup.R hyperedge suggests altered limbic–olfactory organization in PD (Georgiopoulos et al., 2019). In Fig. 9(b), the AMYG.L–PHG.L–AMYG.R–TPOmid.L–PCG.L hyperedge indicates differences in amygdala-

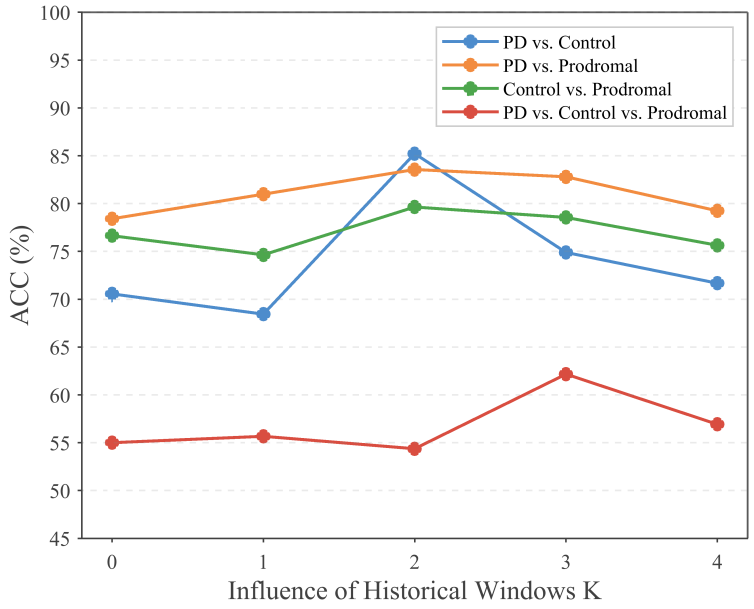

Figure 7: Effect of history length on PPMI.

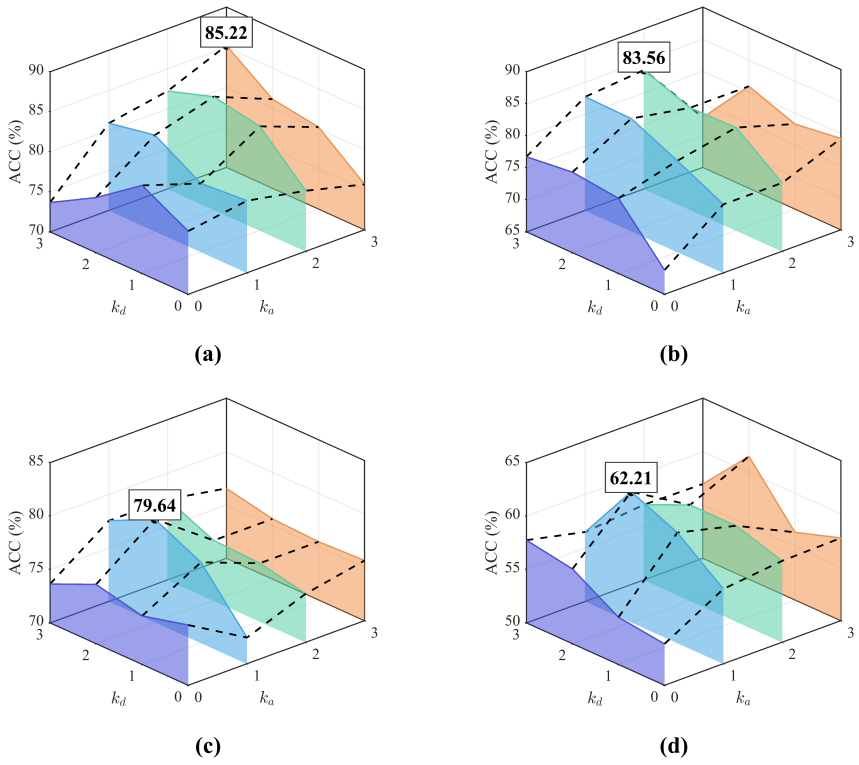

Figure 8: Effect of hypergraph topology editing scores on classification performance. (a) PD vs. HC, (b) PD vs. Prodromal, (c) HC vs. Prodromal, and (d) PD vs. HC vs. Prodromal.

centered limbic and temporal organization between PD and prodromal participants (Wang et al., 2023a). In Fig. 9(c), the AMYG.L–HES.R–AMYG.R–OLF.L–PCG.L hyperedge suggests that olfactory–limbic and auditory organization contributes to distinguishing HC from prodromal participants, broadly consistent with olfactory dysfunction and network dysconnectivity reported during prodromal PD (Liu et al., 2017; Dayan & Browner, 2017). In Fig. 9(d), the AMYG.L–PHG.L–

TPOsup.L–OLF.L–AMYG.R hyperedge highlights a shared olfactory–limbic–temporal pattern that may jointly discriminate HC, prodromal, and PD states (Liu et al., 2017; Georgiopoulos et al., 2019).

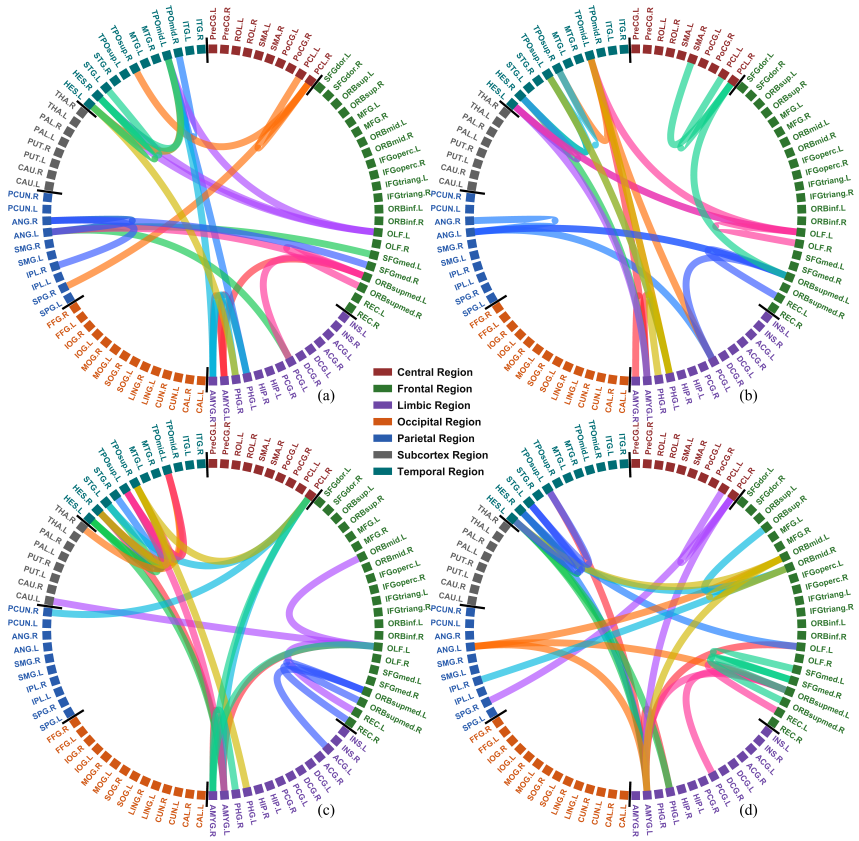

Figure 9: Top-10 discriminative hyperedges on PPMI.(a) PD vs. HC, (b) PD vs. Prodromal,(c) HC vs. Prodromal, and(d) PD vs. HC vs. Prodromal.

Discriminative ROIs. Figure 10 presents the top-10 ROIs ranked by class-wise node–hyperedge membership differences. In Fig. 10(a), AMYG.L, HES.L, ORBsupmed.L, and TPOmid.L rank highly, consistent with reported abnormalities in amygdala, auditory, temporal, and temporo-amygdala-orbitofrontal networks in early PD (Zeng et al., 2022). In Fig. 10(b), the bilateral amygdalae and Heschl's gyri, together with ORBsupmed.L and PCG.L, highlight limbic and temporal differences between PD and prodromal participants (Wang et al., 2023a). In Fig. 10(c), the bilateral amygdalae, orbitofrontal cortex, and temporal poles rank highly, suggesting that olfactory–limbic and temporal systems provide discriminative information between HC and prodromal participants (Liu et al., 2017; Dayan & Browner, 2017). In Fig. 10(d), OLF.L, the bilateral amygdalae and Heschl's gyri, and ORBsupmed.L are prominent, indicating complementary olfactory–limbic and auditory–temporal contributions to three-class discrimination (Liu et al., 2017; Georgiopoulos et al., 2019).

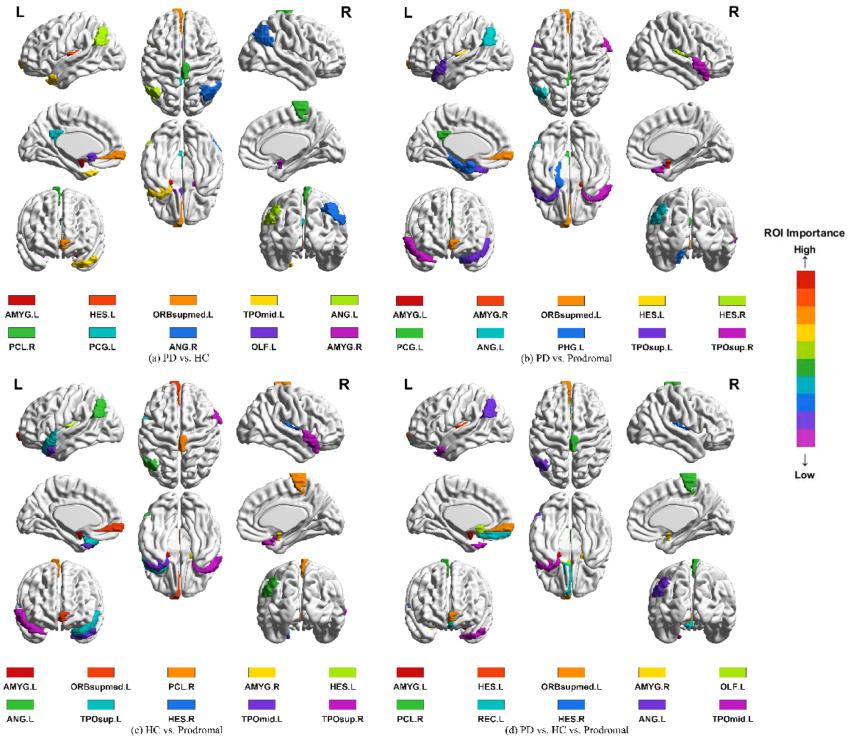

Figure 10: **Top-10 discriminative ROIs on PPMI ranked by node–hyperedge membership differences.**