BRAINIPAGENT: MULTI-AGENT INTERPRETABLE STRUCTURAL PRIORS FOR MULTIMODAL BRAIN NETWORK LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

The effective integration of structural connectivity (SC) and functional connectivity (FC) is crucial for brain disorder diagnosis. However, existing methods remain predominantly data-driven and fail to effectively incorporate interpretable priors, leading to limited circuit-level interpretability. Static multimodal fusion strategies further limit the adaptive reliance on structural priors. In this paper, we propose BrainIPAgent, a multimodal brain network learning framework with multi-agent interpretable structural priors and RL-guided alignment. BrainIPAgent maps structural patterns into an attention bias in the logit space, enabling soft control over information flow in functional representation learning. To preserve structural consistency across hierarchies, we introduce a mask propagation operator that projects structural constraints through soft pooling. Multimodal fusion is formulated as a bilevel optimization problem, where contrastive alignment models the distributional discrepancy between textual embeddings derived from iteratively refined clinical reports and continuous functional representations. An RL-guided controller then adaptively regulates the strengths of structural guidance and contrastive alignment throughout training. Experiments on ADNI and real-world datasets demonstrate that BrainIPAgent achieves superior classification performance while providing interpretable insights. The code is available at `https://anonymous.4open.science/r/BrainIPAgent`.

1 INTRODUCTION

The diagnosis of brain disorders increasingly relies on the holistic integration of multimodal brain connectivity patterns (Lynn & Bassett, 2019; Luo et al., 2024). Structural connectivity (SC), derived from diffusion magnetic resonance imaging (dMRI), characterizes anatomical scaffolding, while functional connectivity (FC), estimated from resting-state functional MRI (rs-fMRI), reflects dynamic neural coordination (Park & Friston, 2013; Hou et al., 2026). Previous studies have shown that abnormalities in structural and functional brain connectivity reflect meaningful alterations associated with brain disorders (Kan et al., 2022; Peng et al., 2025a; Xue et al., 2025). Joint consideration of structural and functional brain connectivity provides more informative characterizations of brain organization than either modality alone (Fotiadis et al., 2024). Motivated by this insight, recent works have further explored multimodal learning frameworks that integrate structural and functional brain connectivity, demonstrating the effectiveness of such joint modeling for brain disorder analysis (Yang et al., 2024; 2026; Wang et al., 2025).

Recent studies have sought to improve the interpretability of brain connectivity models by identifying important structural connections (Li et al., 2023) or extracting discriminative subgraphs (Luo et al., 2025) that contribute most to classification decisions. Concurrently, emerging evidence suggests that brain disorders disrupt not only pairwise connections but also higher-order interactions among groups of brain regions, highlighting the need to model complex synergy patterns (Yu et al., 2024; Qiu et al., 2024b; Abid et al., 2025). Despite these advances, identifying such patterns remains fundamentally data-driven, deriving importance scores purely from model sensitivity rather than domain-grounded reasoning. To overcome this, subsequent works have explored incorporating functional and structural priors into brain network modeling (Li et al., 2025; Pei et al., 2025). However, these priors are typically predefined and restricted to static, pairwise topological constraints.

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.

This rigidity limits the modeling of dynamic, higher-order brain organization and hinders the explicit reasoning essential for clinical diagnosis.

Furthermore, addressing the inherent heterogeneity between structural and functional modalities remains a challenge for existing frameworks (Xia et al., 2025). Current strategies typically rely on static concatenation or fixed attention, failing to capture the evolving relationship between anatomical scaffolding and functional dynamics (Feng et al., 2025; Wang et al., 2025). Such static treatment of discrete SC priors and continuous FC representations risks modality collapse (Yang et al., 2024; Zhou et al., 2025). Crucially, the optimal balance is dynamic, where structural constraints facilitate early alignment but restrict plasticity in later stages. The absence of phase-dependent modulation in existing approaches thus hinders effective convergence and generalization across clinical datasets.

To this end, we propose the multi-agent interpretable structural priors and RL-guided alignment framework (BrainIPAgent) for multimodal brain network learning. As illustrated in Figure 1, BrainIPAgent employs a hierarchical multi-agent framework driven by large language models (LLMs) that integrates interpreter, clinician, and expert agents to emulate clinical reasoning. The multi-agent framework extracts informative pairwise and high-order structural patterns and generates clinical reports refined through iterative expert review. BrainIPAgent utilizes these extracted patterns to construct attention biases in the logit space and employs a mask propagation operator to project these constraints through hierarchical soft pooling. Simultaneously, it enriches functional representations via textual embeddings derived from the clinical reports. Furthermore, we formulate multimodal fusion as a bilevel optimization problem to reconcile the inherent distributional discrepancy between these discrete textual structural embeddings and continuous functional representations. A reinforcement learning (RL) controller is introduced to adaptively regulate the strengths of structural guidance and contrastive alignment throughout training.

The major contributions of this work are listed below:

- We propose a hierarchical multi-agent framework comprising interpreter, clinician, and expert agents that mimics collaborative and iterative clinical reasoning. This framework extracts informative pairwise and high-order structural patterns to construct domain-aware textual priors and sparsity masks, characterizing brain organization at the circuit level.
- We develop a structure-aware attention mechanism that maps extracted patterns to attention biases in the logit space. To ensure anatomical consistency across network scales, we introduce a mask propagation operator that projects these structural constraints through hierarchical soft pooling.
- We design a reinforcement learning controller to bridge the distributional gap between discrete textual priors and continuous functional representations. The controller harmonizes rigid anatomical scaffolding with evolving functional plasticity by dynamically regulating structural guidance and alignment weights.

2 Related Works

Interpretability of Brain Network Existing studies have introduced various interpretability mechanisms in brain network analysis (Cui et al., 2023). IBGNN (Cui et al., 2022) learned global edge masks to pinpoint disorder-specific subgraphs. More recently, attention-based models, such as GGT (Qu et al., 2024) and BISTformer (Li et al., 2024), leverage self-attention weights to highlight salient functional connections. Meanwhile, hierarchical methods, including MAHGCN (Liu et al., 2024) and HSGPL (Tang et al., 2024a), employ multiscale atlases or pooling strategies to identify biomarkers at different organizational levels. However, these methods rely on statistical correlations, failing to elucidate the underlying clinical rationale. BrainIPAgent leverages a multi-agent framework to derive antecedent clinical priors for domain-grounded interpretation.

Multi-Agent Frameworks in Healthcare Multi-agent frameworks have revolutionized clinical automation. MedAgents (Tang et al., 2024b) and MDAgents (Kim et al., 2024) simulate medical consultation scenarios to enhance decision-making through collaborative debate. For medical imaging, MedRAX (Fallahpour et al., 2025) and LungNoduleAgent (Yang et al., 2025a) leverage agent-based interactions to generate comprehensive chest X-ray reports and precise nodule diagnoses. To optimize collaboration efficiency, ProAgent (Zhang et al., 2024) introduces proactive mechanisms for

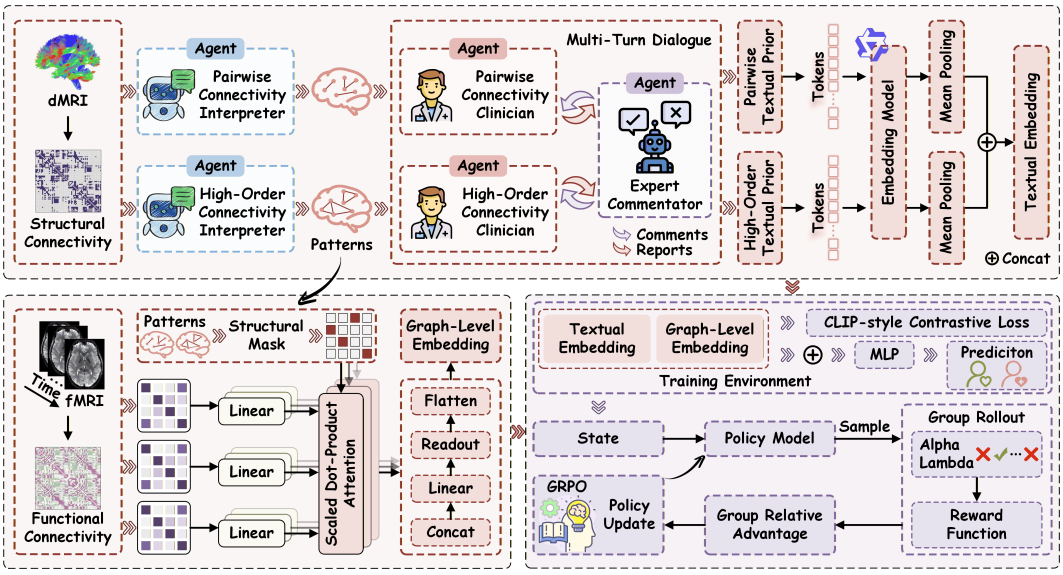


Figure 1: Overview of the BrainIPAgent framework. (1) The **Multi-Agent Interpretation Framework** extracts structural patterns via collaborative reasoning to generate sparsity masks and textual priors. (2) The **Structure-Aware Functional Encoder** maps these patterns to attention biases in the logit space and projects them via mask propagation. (3) The **RL-Guided Alignment Module** dynamically regulates structural guidance strength $Alpha$ (α) and alignment weight $Lambda$ (λ).

adaptive coordination. Furthermore, MedDQA (Peng et al., 2025b) integrates multi-source data for medical question answering. While multi-agent frameworks are promising in healthcare, their application in brain network analysis remains unexplored. BrainIPAgent converts multi-agent clinical reasoning into structure-aware attention masks and comprehensive clinical reports.

Multimodal Fusion in Brain Network Analysis Multimodal fusion has emerged as a promising paradigm. Foundational frameworks like GBDM (Zhang et al., 2021) established joint embedding spaces to synthesize heterogeneous data, while MMP-GCN (Song et al., 2023) incorporated structural constraints via multicenter pooling. HKC (Yang et al., 2025b), ATPGCN (Bian et al., 2024), and TMHGC-CBNF (Feng et al., 2026) integrate hyperbolic manifolds, persistent homology, or hypergraphs to capture complex topological dependencies across modalities. Similarly, biologically inspired models like SCNNs (Wei et al., 2025) and MSSTAN (Kong et al., 2025) explore spike-based coupling mechanisms and multi-scale attention to refine dynamic representations. Nevertheless, these static approaches overlook the discrepancy between anatomical scaffolding and functional dynamics. BrainIPAgent employs a reinforcement learning controller for adaptive alignment.

3 METHODS

We construct subject-level SC and FC from dMRI and rs-fMRI data, respectively. Given a multimodal brain network dataset $\{\mathrm{X}_n\}_{n=1}^{N}$ with N subjects, each sample consists of two modalities. Specifically, let $\mathbf{A}^{\mathrm{S}} \in \mathbb{R}^{|\Omega|\times|\Omega|}$ and $\mathbf{A}^{\mathrm{F}} \in \mathbb{R}^{|\Omega|\times|\Omega|}$ denote the structural and functional connectivity matrices of a subject, where $|\Omega| = 90$ is the number of brain regions. The multimodal representation of a subject is thus given by $\mathrm{X} = \{\mathbf{A}^{\mathrm{S}}, \mathbf{A}^{\mathrm{F}}\}$. The goal is to learn a decision function $f : \mathrm{X} \to y$ that maps multimodal connectivity data to a subject-level label y (e.g., patient or normal control).

3.1 MULTI-AGENT INTERPRETATION FRAMEWORK FOR STRUCTURAL PRIOR

We propose a subject-level multi-agent interpretation framework to construct a textual prior $\mathcal{T}$ from structural connectivity (SC). The Pairwise Connectivity Interpreter Agent (PCI Agent) and the High-Order Connectivity Interpreter Agent (HCI Agent) first extract informative pairwise and higher-order structural patterns $\mathcal{P}_{\mathrm{p}}$ and $\mathcal{P}_{\mathrm{h}}$ from the raw connectome. The Pairwise Connectivity

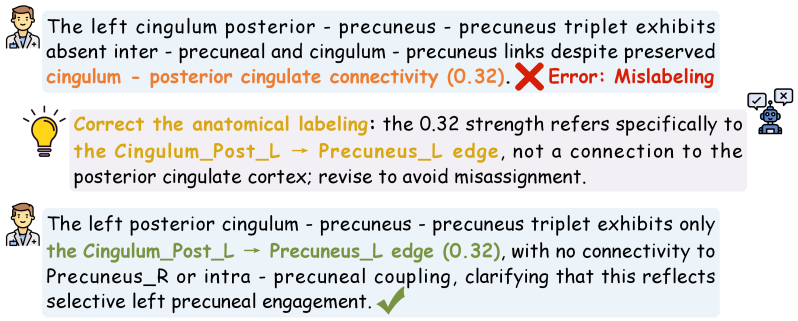


Figure 2: Illustration of multi-turn dialogue between the EC Agent and the Clinician Agent. The EC Agent identifies and corrects anatomical labeling errors to refine the clinical report.

Clinician Agent (PCC Agent) and the High-Order Connectivity Clinician Agent (HCC Agent) then convert these selected patterns into explanatory clinical reports. The Expert Commentator Agent (EC Agent) further refines these reports through multi-turn dialogue to improve conceptual clarity and anatomical-functional consistency. After T iterations, the finalized reports $\mathcal{R}_{\mathrm{p}}^{T}$ and $\mathcal{R}_{\mathrm{h}}^{T}$ form domain-aware textual priors that are encoded and integrated into the predictive model.

Connectivity Interpreter Agents. Two interpreter agents are employed to operate on the subject-level SC matrix $\mathbf{A}^{\mathrm{S}}$, capturing structural information at both pairwise and high-order levels. The PCI Agent identifies structurally meaningful direct links by applying anatomically informed constraints rather than simple magnitude ranking. It selects the top D pairwise structural patterns, denoted as

$$\mathcal{P}_{\mathrm{p}} = \{(r_i, r_j, w_{ij})_d\}_{d=1}^{D}, \tag{1}$$

where r_i and r_j represent the indices of the brain regions involved in the d-th pair, with $1 \leq i, j \leq |\Omega|$, and $w_{ij} = \mathbf{A}_{ij}^{\mathrm{S}}$ denotes the structural connectivity strength.

The HCI Agent further characterizes circuit-level organization by searching for triplet-based co-connectivity configurations. It identifies the top D higher-order structural patterns, denoted as

$$\mathcal{P}_{\mathrm{h}} = \{(r_i, r_j, r_k, w_{ij}, w_{ik}, w_{jk})_d\}_{d=1}^{D}, \tag{2}$$

where r_i, r_j, and r_k represent the indices of the brain regions involved in the d-th triplet, with $1 \leq i, j, k \leq |\Omega|$, and w_{ij}, w_{ik}, w_{jk} denote the structural connectivity strengths.

These extracted pattern sets $\mathcal{P}_{\mathrm{p}}$ and $\mathcal{P}_{\mathrm{h}}$ serve as clinically meaningful prompts, transforming raw numerical data into structured input for the subsequent text generation stage.

Connectivity Clinician Agents. We employ two clinician agents to translate the identified patterns into concise, clinically oriented explanatory reports. The PCC Agent processes the pairwise set $\mathcal{P}_{\mathrm{p}}$ to emphasize the anatomical identity of involved regions and their neuroanatomical significance. In parallel, the HCC Agent interprets the higher-order set $\mathcal{P}_{\mathrm{h}}$ to offer insights into the topological organization of brain circuits involved in complex processing. Formally, the generation of these clinical reports is defined as

$$\mathcal{R}_{\mathrm{p}} = \Phi_{\mathrm{LLM}}(\mathrm{Prompt}(\mathcal{P}_{\mathrm{p}})), \quad \mathcal{R}_{\mathrm{h}} = \Phi_{\mathrm{LLM}}(\mathrm{Prompt}(\mathcal{P}_{\mathrm{h}})), \tag{3}$$

where Φ_{LLM} represents the large language model function that generates the textual report, and $\mathrm{Prompt}(\cdot)$ refers to the process of formatting the patterns into a structured prompt.

Expert Commentator Agent. The EC Agent evaluates the previous reports $\mathcal{R}_{\mathrm{p}}^{t-1}$ and $\mathcal{R}_{\mathrm{h}}^{t-1}$ with respect to neuroanatomical plausibility and conceptual clarity, and outputs structured revision suggestions. These suggestions are then fed back to the clinician agents, which are responsible for incorporating the feedback and generating the updated reports in the next iteration. At iteration t, the refinement process is

$$\mathcal{R}_{\mathrm{p}}^{t}, \mathcal{R}_{\mathrm{h}}^{t} = \Phi_{\mathrm{LLM}}\left(\mathrm{Prompt}\left(\mathcal{R}_{\mathrm{p}}^{t-1}, \mathcal{R}_{\mathrm{h}}^{t-1}, \mathcal{C}^{t}\right)\right), \tag{4}$$

where $\mathcal{C}^{t}$ denotes the expert commentary provided by the EC Agent. An example of the multi-turn refinement process is shown in Figure 2, where the EC Agent detects anatomical labeling errors and guides the clinician agent to revise the report.

This feedback-driven interaction is repeated for a fixed number of rounds T. The final reports $\mathcal{R}_{\mathrm{p}} = \mathcal{R}_{\mathrm{p}}^{T}$ and $\mathcal{R}_{\mathrm{h}} = \mathcal{R}_{\mathrm{h}}^{T}$ are aggregated as the subject-level textual prior $\mathcal{T}$. Each report is encoded by Qwen3-Embedding-8B (Zhang et al., 2025) to obtain $\mathbf{Z}^{\mathrm{p}} = \Phi_{\mathrm{emb}}(\mathcal{R}_{\mathrm{p}})$ and $\mathbf{Z}^{\mathrm{h}} = \Phi_{\mathrm{emb}}(\mathcal{R}_{\mathrm{h}})$. These embeddings are concatenated to form the final textual embedding $\mathbf{Z}^{\mathcal{T}} = [\mathbf{Z}^{\mathrm{p}} \,\|\, \mathbf{Z}^{\mathrm{h}}]$.

3.2 Structure-Aware Functional Encoder

We utilize the pairwise $\mathcal{P}_{\mathrm{p}}$ and higher-order $\mathcal{P}_{\mathrm{h}}$ patterns to construct a structure-aware mask, employing hierarchical pooling and mask propagation to capture multi-scale functional organization while guiding the encoder toward neurobiologically meaningful interactions supported by SC.

Structure-Aware Attention Mask. To inject structural guidance into functional modeling, we construct a sparse structure-aware mask $\mathbf{M} \in \mathbb{R}^{|\Omega|\times|\Omega|}$ from the structural pattern sets P_p and P_h. Specifically, we construct a symmetric mask by activating entries corresponding to each $(r_i, r_j) \in P_p$ and the constituent pairs of each $(r_i, r_j, r_k) \in P_h$.

The mask is integrated into the attention computation as an additive bias to the attention logits, with its contribution adaptively scaled by α to enhance the model's focus on structurally relevant interactions. We generate structure-aware node embedding $\mathbf{Z}^{L}$ through multi-head self-attention, with L-layer nonlinear mapping and M attention heads. For layer l, the output $\mathbf{Z}^{l}$ is computed as

$$\mathbf{Z}^{l} = \left(\|_{m=1}^{M}\, h^{l,m}\right) \mathbf{W}_{\mathcal{O}}^{l}, \tag{5}$$

$$h^{l,m} = \mathrm{Softmax}\left(\frac{\left(\mathbf{Z}^{l-1}\mathbf{W}_{\mathcal{Q}}^{l,m}\right)\left(\mathbf{Z}^{l-1}\mathbf{W}_{\mathcal{K}}^{l,m}\right)^{\top}}{\sqrt{d_K^{l,m}}} + \alpha\mathbf{M}^{l}\right)\left(\mathbf{Z}^{l-1}\mathbf{W}_{\mathcal{V}}^{l,m}\right), \tag{6}$$

where $\|$ denotes the concatenation operator, l and m denote the layer index and the head index, $\mathbf{W}_{\mathcal{O}}^{l}$, $\mathbf{W}_{\mathcal{Q}}^{l,m}$, $\mathbf{W}_{\mathcal{K}}^{l,m}$, $\mathbf{W}_{\mathcal{V}}^{l,m}$ are learnable projection matrices.

Hierarchical Pooling and Mask Propagation. A hierarchical pooling strategy is adopted to capture multi-scale functional organization, where node representations are progressively aggregated into coarser abstractions through a soft assignment mechanism. Structural consistency across hierarchical levels is maintained by propagating the mask through the pooling layers as follows:

$$\mathbf{M}^{l} = \left(\mathbf{P}^{l-1}\right)^{\top}\mathbf{M}^{l-1}\mathbf{P}^{l-1}, \quad \mathbf{P}_{ic} = \frac{e^{\langle \mathbf{Z}_i^{l}, \mathbf{E}_c\rangle}}{\sum_{c'}^{C} e^{\langle \mathbf{Z}_i^{l}, \mathbf{E}_{c'}\rangle}}, \tag{7}$$

where $\mathbf{P}^{l-1} \in \mathbb{R}^{N_{l-1}\times N_l}$ denotes the soft assignment matrix at layer $l-1$, which probabilistically assigns node i to cluster c, and $\mathbf{E}_c$ denotes the cluster centers, initialized using Xavier uniform initialization followed by the Gram-Schmidt process to obtain orthonormal bases.

After L layers of hierarchical encoding, we obtain the final node embeddings $\mathbf{Z}^{L}$. To obtain the graph-level embedding, we apply a readout function that aggregates $\mathbf{Z}^{L}$ using the soft assignment matrix $\mathbf{P}$. The graph-level embedding is derived as $\mathbf{Z}^{\mathrm{G}} = \mathbf{P}^{\top}\mathbf{Z}^{L}$.

3.3 RL-Guided CLIP-style Representation Alignment

While structural guidance is incorporated into the functional encoder, a fixed guidance strength limits the flexibility to adjust reliance on priors across training stages. To address this, we introduce an RL-guided mechanism that dynamically regulates the coupling between functional representations and textual priors. Specifically, we employ a CLIP-style contrastive objective for semantic consistency and an RL controller to adaptively modulate structural guidance strength α and alignment weight λ.

3.3.1 CLIP-style Contrastive Representation Alignment

The functional representation $\mathbf{Z}^{\mathrm{G}}$ and the structural textual prior $\mathbf{Z}^{\mathcal{T}}$ are projected into a shared embedding space via separate projection heads, yielding projected embeddings $\mathbf{H}^{\mathrm{G}}$ and $\mathbf{H}^{\mathcal{T}}$. After ℓ_2-normalization, we compute the symmetric InfoNCE loss to enforce cross-modal consistency:

$$\mathcal{L}_{\mathrm{CL}} = \frac{1}{2}\left(\mathcal{L}_{\mathrm{InfoNCE}}(\mathbf{H}^{\mathrm{G}}, \mathbf{H}^{\mathcal{T}}) + \mathcal{L}_{\mathrm{InfoNCE}}(\mathbf{H}^{\mathcal{T}}, \mathbf{H}^{\mathrm{G}})\right), \tag{8}$$

where $\mathcal{L}_{\text{InfoNCE}}(\mathbf{U}, \mathbf{V})$ is the contrastive objective:

$$\mathcal{L}_{\text{InfoNCE}}(\mathbf{U}, \mathbf{V}) = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\exp\left(\langle \mathbf{u}_i, \mathbf{v}_i \rangle / \tau\right)}{\sum_{j=1}^{B} \exp\left(\langle \mathbf{u}_i, \mathbf{v}_j \rangle / \tau\right)}, \tag{9}$$

where B is the batch size and τ is a temperature parameter.

To leverage the complementary information from both modalities for diagnosis, we fuse the original representations $\mathbf{Z}^{\text{G}}$ and $\mathbf{Z}^{\mathcal{T}}$ to derive the subject-level prediction:

$$\hat{\mathbf{y}} = \text{Softmax}\left(\mathcal{F}_{\text{cls}}\left(\mathbf{Z}^{\text{G}} \parallel \mathbf{Z}^{\mathcal{T}}\right)\right), \tag{10}$$

where $\mathcal{F}_{\text{cls}}$ denotes the classification head.

3.3.2 Adaptive Structure-Aware Alignment with Reinforcement Learning

The RL module functions as a high-level controller that adaptively regulates the structural guidance strength α and the contrastive alignment weight λ. At each training epoch e, the controller observes a global state $\mathbf{s}_e$ summarizing training dynamics and outputs an action (α_e, λ_e) applied consistently throughout the subsequent epoch, decoupling control dynamics from weight updates.

We optimize the control policy using a modified group relative policy optimization (GRPO) strategy under non-stationary training feedback (Shao et al., 2024). Given the state $\mathbf{s}_e$, the controller samples K candidate actions $\{(\alpha_e^k, \lambda_e^k)\}_{k=1}^{K}$ and evaluates each through a trial phase to obtain the reward r_e^k:

$$r_e^k = \underbrace{\left(\text{Acc}_{\text{trial}}^k - \text{Acc}_{\text{prev}}\right)}_{\text{Performance Gain}} - \underbrace{\beta_1 \mathcal{H}\left(\mathbf{S}_{\text{attn}}^k\right)}_{\text{Attention Sparsity}} - \underbrace{\beta_2 \mathcal{H}\left(\mathbf{P}^k\right)}_{\text{Cluster Certainty}}, \tag{11}$$

where $\text{Acc}_{\text{trial}}^k$ and Acc_{prev} denote the validation accuracy of the k-th candidate trial and the previous epoch, respectively, $\mathcal{H}(\cdot)$ denotes the entropy function, and $\mathbf{S}_{\text{attn}}^k$ and $\mathbf{P}^k$ represent the attention weight matrices and soft assignment probabilities of the k-th candidate trial.

To mitigate gradient variance without introducing instability associated with variance estimation in small groups, we compute the advantage by centering the rewards with respect to the group mean, omitting the standard deviation normalization term:

$$A_e^k = r_e^k - \frac{1}{K}\sum_{j=1}^{K} r_e^j. \tag{12}$$

The policy parameters are updated to maximize the expected advantage, encouraging the selection of control actions that yield higher group-relative rewards.

The overall training objective integrates supervised learning with adaptive multimodal regularization. The base network is optimized under control actions (α_e, λ_e) by minimizing the total objective:

$$\mathcal{L}_e = \mathcal{L}_{\text{CE}} + \lambda_e \mathcal{L}_{\text{CL}}, \tag{13}$$

where $\mathcal{L}_{\text{CE}}$ denotes the cross-entropy loss for subject-level prediction and $\mathcal{L}_{\text{CL}}$ is the CLIP-style contrastive loss. The training process follows a bilevel optimization scheme. At the inner level, the base network parameters are updated using Adam optimization. At the outer level, the RL controller is optimized using validation feedback. This separation enables adaptive regulation of structural guidance and representation alignment.

4 Experiments

4.1 Experimental Settings

Datasets. We used dMRI and rs-fMRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Jack et al., 2008) and two clinical sites. The ADNI cohort included 225 normal controls (NCs) and 203 patients with mild cognitive impairment (MCI). The clinical cohorts included 61 patients with major depressive disorder (MDD) and 47 NCs from the Affiliated Zhongda Hospital

Table 1: Performance comparison with state-of-the-art methods across four evaluation settings. We report the classification accuracy (Mean ± Std). The best results are highlighted in **bold**, and the second-best are underlined. Comprehensive results are provided in **Appendix D**.

Method	Modality	Datasets			
		ADNI	Zhongda	Xinxiang	ZDXX
BrainNetTF (Kan et al., 2022)	FC	71.54 ± 6.17	74.87 ± 8.72	83.63 ± 12.18	70.82 ± 8.13
BrainNetTF (Kan et al., 2022)	SC	67.32 ± 7.93	75.97 ± 8.13	83.31 ± 8.08	70.06 ± 8.30
ALTER (Yu et al., 2024)	FC	71.20 ± 7.70	74.14 ± 5.14	80.64 ± 8.76	72.26 ± 7.24
ALTER (Yu et al., 2024)	SC	68.82 ± 9.80	75.05 ± 6.91	79.52 ± 7.31	69.09 ± 8.79
AL-NEGAT (Chen et al., 2024)	FC, SC	71.73 ± 6.21	78.70 ± 4.03	88.53 ± 10.90	77.76 ± 7.16
NeuroPath (Wei et al., 2024)	FC, SC	72.70 ± 5.47	72.32 ± 8.25	83.13 ± 14.40	69.65 ± 10.57
MS-Inter-GCN (Xia et al., 2025)	FC, SC	74.62 ± 8.88	77.77 ± 7.25	84.74 ± 11.03	71.18 ± 10.78
ST2GCN (Li et al., 2025)	FC, SC	74.31 ± 4.26	78.59 ± 6.25	88.67 ± 8.44	77.62 ± 7.41
HKGCN (Yang et al., 2026)	FC, SC	73.86 ± 5.58	75.96 ± 4.24	84.56 ± 9.48	75.42 ± 3.29
HKGAT (Yang et al., 2026)	FC, SC	76.83 ± 2.96	76.87 ± 4.42	85.99 ± 8.40	75.98 ± 6.70
BrainIPAgent (ours)	FC, SC	**79.94 ± 5.81**	**82.33 ± 5.07**	**91.35 ± 7.58**	**80.61 ± 7.16**

Table 2: Ablation study results on the ADNI and ZDXX datasets. We report the classification accuracy (Mean ± Std). The best results are highlighted in **bold**, and the second-best are underlined.

Model	Components			Datasets	
	Textual Alignment	Mask	RL Adaptation	ADNI	ZDXX
Baseline (Naive Fusion)	×	×	×	74.95 ± 6.08	76.76 ± 5.45
w/o Textual Alignment	×	✓	✓	76.12 ± 6.89	77.45 ± 5.98
w/o Structural Guidance	✓	×	✓	76.95 ± 6.17	78.06 ± 4.38
w/o RL Adaptation	✓	✓	×	77.12 ± 5.03	78.12 ± 4.86
Ours (Full Model)	✓	✓	✓	**79.94 ± 5.81**	**80.61 ± 7.16**

of Southeast University (Zhongda), and 31 MDD patients and 46 NCs from the Second Affiliated Hospital of Xinxiang Medical University (Xinxiang). All participants provided informed consent according to the protocols of their respective studies. We further combined the two clinical cohorts into a multi-site dataset (ZDXX). Detailed preprocessing pipelines are provided in **Appendix A**.

Implementation Details. The proposed method was trained for 500 epochs with a batch size of 100, employing the Adam optimizer with a learning rate of 5×10^{-5}. The multi-agent iteration was set to $T = 5$. In the GRPO implementation, the group size of candidate actions was set to $K = 4$. A 10-fold cross-validation strategy was employed to ensure robust evaluation. All experiments were implemented in PyTorch and trained on two NVIDIA RTX 4090 GPUs. Additional implementation details of the data split, RL controller, and training configuration, together with detailed prompt templates, are provided in **Appendix B** and **Appendix I**, respectively.

4.2 Experimental Results

We compare our method against eight baselines, spanning single-modal and multimodal approaches. Single-modal methods **BrainNetTF** (Kan et al., 2022) and **ALTER** (Yu et al., 2024) model complex topological structures. Among multimodal methods, **AL-NEGAT** (Chen et al., 2024) employs adversarial learning, **NeuroPath** (Wei et al., 2024) adopts biological pathway Transformers, **MS-Inter-GCN** (Xia et al., 2025) and **ST2GCN** (Li et al., 2025) focus on spatio-temporal modeling, and **HKGCN** alongside **HKGAT** (Yang et al., 2026) leverage hyperbolic geometry.

The quantitative results across four evaluation settings are presented in Table 1. BrainIPAgent consistently achieves state-of-the-art performance, outperforming all single-modal and multimodal baselines. Specifically, our method surpasses strong competitors such as ST2GCN, AL-NEGAT, and HKGAT by margins ranging from approximately 2.68% to 3.63% in accuracy. Notably, on the ADNI and ZDXX datasets, where baseline performance varies considerably, BrainIPAgent maintains a clear lead. This demonstrates that by leveraging SC as a prior to model complex topological interactions, our method achieves robust and consistent performance across diverse data sources.

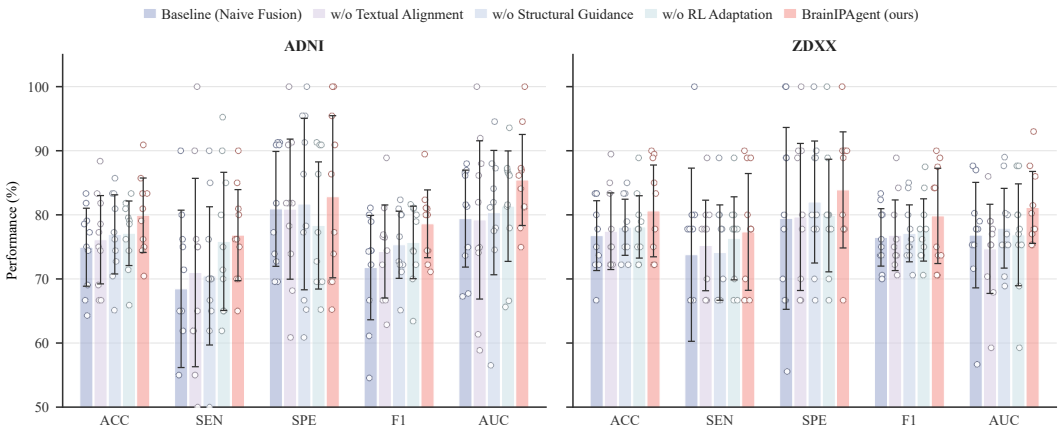
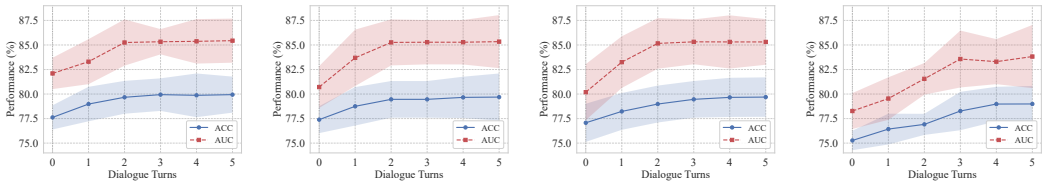


Figure 3: Ablation study results on ADNI and ZDXX. Bars and error bars indicate Mean ± Std, and dots denote results from individual folds.



(a) Qwen-3.5 (b) Gemini-3 (c) DeepSeek-V4 (d) Qwen-3.5-9B

Figure 4: Performance evolution on the ADNI dataset over dialogue turns ($T = 0 \rightarrow 5$) for (a) **Qwen-3.5**, (b) **Gemini-3**, (c) **DeepSeek-V4**, and (d) **Qwen-3.5-9B**. Blue circles and red squares denote Accuracy and AUC, respectively. Shaded bands indicate the Standard Error of the Mean.

4.3 Ablation Study

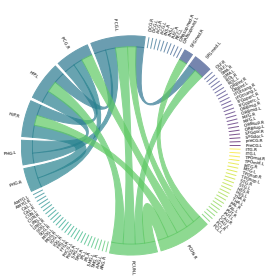
We assess the contribution of each key component on ADNI and ZDXX. Table 2 provides a concise summary of the ablation study, while Fig. 3 presents the detailed comparisons across all metrics and individual folds. Removing textual alignment causes a clear performance drop, indicating that textual priors provide complementary semantic information beyond structural guidance. Removing structural guidance also consistently degrades performance, confirming the effectiveness of the structure-aware mask in promoting biologically plausible functional interactions. Finally, fixing α and λ reduces performance compared with the full model, demonstrating the benefit of adaptively regulating structural and semantic priors across training stages.

4.4 Analysis of LLM Backbones and Multi-Turn Dialogue

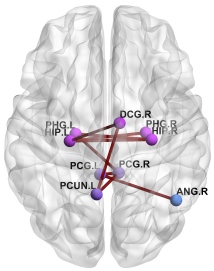
We investigate the effects of LLM backbones and multi-turn refinement using four models: **Qwen-3.5** (Team, 2026), **Gemini-3** (Team et al., 2023), **DeepSeek-V4** (DeepSeek-AI et al., 2026), and the lightweight **Qwen-3.5-9B** (Team, 2026). As shown in Figure 4, multi-turn interaction with the EC Agent consistently improves performance over the initial generation ($T = 0$), with most gains occurring between $T = 0$ and $T = 2$ and diminishing thereafter. This suggests that iterative collaboration refines early reasoning and converges efficiently within a few dialogue turns. The framework remains robust across LLM backbones, including Qwen-3.5-9B, which benefits from refinement, indicating that agent collaboration improves reasoning quality even for parameter-constrained models.

4.5 Interpretability and Analysis

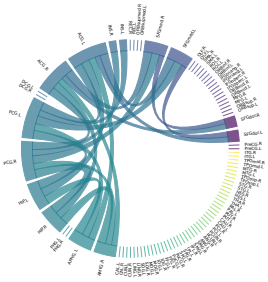
We visualize discriminative topological features across subjects and representative clinical interpretation reports from an MCI patient to assess the interpretability of BrainIPAgent in Figure 5.



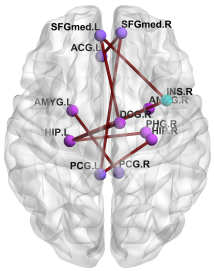
(a) Prior (ADNI)



(b) Attention (ADNI)



(c) Prior (ZDXX)



(d) Attention (ZDXX)

Pairwise Connectivity Clinician Agent: The profile reveals **asymmetric disruptions** in hippocampal network integration. Bilateral hippocampal-parahippocampal connections remain **preserved (0.36, 0.33)**, supporting intact memory encoding circuitry. However, key **left-sided pathways** linking the hippocampus to the cingulum, insula, and precuneus are **completely disconnected**. This **zero-strength connectivity** indicates a coordinated breakdown across default mode and salience networks.

(e) Pairwise connectivity clinician report

High-Order Connectivity Clinician Agent: This profile reveals early, **left-lateralized disintegration** of default mode hubs contrasted with **right-hemisphere preservation**. Specifically, the left anterior cingulate-hippocampal-precuneus circuit **lacks all connections (0.00, 0.00, 0.00)**, indicating failed integration. In contrast, the corresponding **right-sided network** demonstrates **preserved interconnectivity** across all three edges (0.33, 0.25, 0.25), reinforcing resilience within the medial temporal-posterior axis.

(f) High-Order connectivity clinician report

Figure 5: Visualization of interpretability results based on Qwen-3.5. (a) and (b) present the results on the ADNI (NC vs. MCI), while (c) and (d) present the results on the ZDXX (NC vs. MDD). **Prior** displays the top **structural** connections selected by the PCI and HCI Agents, and **Attention** illustrates the top structure-aware **functional** attention weights. (e) and (f) present representative excerpts of the subject-specific clinical interpretation reports generated by the PCC and HCC Agents for an MCI patient. More results are provided in **Appendix H**.

Topological Interpretability. For NC vs. MCI, the agents select structural connections among DMN and medial temporal regions, including HIP, PHG, and PCG (Figure 5(a)), while the structure-aware attention captures corresponding functional synchronization (Figure 5(b)), consistent with established evidence on cognitive decline (Pievani et al., 2014). For NC vs. MDD, the agents identify Cortico-Limbic connections involving SFGmed, ACG, AMYG, and INS (Figure 5(c)), with attention further highlighting disease-specific functional coupling patterns (Figure 5(d)) (Drysdale et al., 2017). These results demonstrate the adaptability of the structure-guided design across disorders.

Clinical Interpretation Report Generation. Figures 5(e) and 5(f) display excerpts from clinical interpretation reports for an MCI patient. The PCC Agent identifies asymmetric disruptions, where key left-sided pathways exhibit zero-valued reconstructed connectivity despite preserved bilateral links. This is consistent with prior evidence of structural and microstructural alterations associated with cognitive impairment and early Alzheimer's disease (Lo et al., 2010; Spotorno et al., 2024; Qiu et al., 2024a). The HCC Agent highlights left-lateralized disintegration, with zero-valued reconstructed connections contrasted by preserved right-sided interconnectivity. These insights demonstrate that BrainIPAgent provides transparent and biologically grounded clinical reasoning.

5 Conclusion

In this paper, we propose BrainIPAgent, a multimodal brain network learning framework with multi-agent interpretable structural priors and RL-guided alignment. BrainIPAgent emulates the iterative reasoning of clinical diagnosis by employing a hierarchical multi-agent system to extract informative structural patterns and domain-aware textual priors. Moreover, a reinforcement controller is introduced to adaptively regulate the strengths of structural guidance and contrastive alignment throughout training. Extensive experiments demonstrate the superior performance and clinically coherent interpretability of BrainIPAgent.

Limitations and Future Work. Primary limitations include the computational inefficiency of fixed-turn interactions and the hallucinations in LLMs. Future work will integrate RL frameworks to design an early-stopping mechanism to minimize inference costs, and employ reinforcement learning from human feedback to enhance the clinical validity of the generated clinical interpretations.

AI USE STATEMENT

Generative AI tools were used in this work in two distinct roles. First, large language models are integral components of the proposed BrainIPAgent framework, where they are employed as multi-agent modules to identify structural connectivity patterns, generate clinical interpretations, and iteratively refine these interpretations through expert feedback. The specific models, prompts, and interaction procedures are described in the main text and **Appendix I**. Second, generative AI tools were used to assist with language polishing, wording refinement, and improving the readability of the manuscript.

All AI-assisted content was reviewed and revised by the authors. The authors verified the scientific claims, mathematical formulations, experimental results, figures, tables, and cited references, and take full responsibility for the final content of the paper. Generative AI tools were not treated as authors and were not used as a substitute for human evaluation of the experimental results or scientific conclusions.

REFERENCES

Shagufta Abid, Dongyu Zhang, Ahsan Shehzad, Jing Ren, Shuo Yu, Hongfei Lin, and Feng Xia. SpeechHGT: A multimodal hypergraph transformer for speech-based early Alzheimer's disease detection. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 9131–9139, 2025.

Chenyuan Bian, Nan Xia, Anmu Xie, Shan Cong, and Qian Dong. Adversarially trained persistent homology based graph convolutional network for disease identification using brain connectivity. *IEEE Transactions on Medical Imaging*, 43(1):503–516, 2024.

Yuzhong Chen, Jiadong Yan, Mingxin Jiang, Tuo Zhang, Zhongbo Zhao, Weihua Zhao, Jian Zheng, Dezhong Yao, Rong Zhang, Keith M. Kendrick, and Xi Jiang. Adversarial learning based node-edge graph attention networks for autism spectrum disorder identification. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6):7275–7286, 2024.

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, 2022.

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, 2023.

DeepSeek-AI, Anyi Xu, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, Chenchen Ling, Chengda Lu, Chenggang Zhao, et al. Deepseek-v4: Towards highly efficient million-token context intelligence, 2026. URL `https://arxiv.org/abs/2606.19348`.

Andrew T Drysdale, Logan Grosenick, Jonathan Downar, Katharine Dunlop, Farrokh Mansouri, Yue Meng, Robert N Fetcho, Benjamin Zebley, Desmond J Oathes, Amit Etkin, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, 23 (1):28–38, 2017.

Adibvafa Fallahpour, Jun Ma, Alif Munim, Hongwei Lyu, and Bo Wang. MedRAX: Medical reasoning agent for chest x-ray. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 15661–15676. PMLR, 2025.

Jingxi Feng, Heming Xu, Junhao Cai, Yujie Chang, Dong Zhang, Shaoyi Du, and Juan Wang. Cross-modal brain graph transformer via function-structure connectivity network for brain disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 247–256. Springer, 2025.

Jingxi Feng, Shaoyi Du, Heming Xu, Rundong Xue, Xiangmin Han, Dong Zhang, Jue Jiang, Yue Gao, and Juan Wang. Topology-aware multi-view hypergraph computation-based cross-modal brain network fusion for brain disease diagnosis. *Information Fusion*, 127:103751, 2026.

Panagiotis Fotiadis, Linden Parkes, Kathryn A Davis, Theodore D Satterthwaite, Russell T Shinohara, and Dani S Bassett. Structure–function coupling in macroscale human brain networks. *Nature Reviews Neuroscience*, 25(10):688–704, 2024.

Tao Hou, Wenhao Dai, Jiashuang Huang, Youyong Kong, and Weiping Ding. Multimodal brain network analysis: Research advances and challenges. *Information Fusion*, 130:104096, 2026.

Clifford R Jack, Jr., Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4): 685–691, 2008.

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.

Yubin Kim, Chanwoo Park, Hyewon Jeong, Yik S Chan, Xuhai Xu, Daniel McDuff, Hyeonhoon Lee, Marzyeh Ghassemi, Cynthia Breazeal, and Hae W Park. MDAgents: An adaptive collaboration of LLMs for medical decision-making. *Advances in Neural Information Processing Systems*, 37:79410–79452, 2024.

Youyong Kong, Xiaotong Zhang, Wenhan Wang, Yue Zhou, Yueying Li, and Yonggui Yuan. Multiscale spatial-temporal attention networks for functional connectome classification. *IEEE Transactions on Medical Imaging*, 44(1):475–488, 2025.

Gaotang Li, Marlena Duda, Xiang Zhang, Danai Koutra, and Yujun Yan. Interpretable sparsification of brain graphs: Better practices and effective designs for graph neural networks. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1223–1234, 2023.

Lanting Li, Liuzeng Zhang, Peng Cao, Jinzhu Yang, Fei Wang, and Osmar R Zaiane. Exploring spatio-temporal interpretable dynamic brain function with transformer for brain disorder diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 195–205. Springer, 2024.

Shengrong Li, Qi Zhu, Chunwei Tian, Wei Shao, and Daoqiang Zhang. Interpretable dynamic brain network analysis with functional and structural priors. *IEEE Transactions on Medical Imaging*, 44(12):4878–4889, 2025.

Mianxin Liu, Han Zhang, Feng Shi, and Dinggang Shen. Hierarchical graph convolutional network built by multiscale atlases for brain disorder diagnosis using functional connectivity. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11):15182–15194, 2024.

Chun-Yi Lo, Pei-Ning Wang, Kun-Hsien Chou, Jinhui Wang, Yong He, and Ching-Po Lin. Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease. *Journal of Neuroscience*, 30(50):16876–16885, 2010.

Xuexiong Luo, Jia Wu, Jian Yang, Shan Xue, Amin Beheshti, Quan Z Sheng, David McAlpine, Paul Sowman, Alexis Giral, and Philip S Yu. Graph neural networks for brain graph learning: A survey. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 8170–8178, 2024.

Xuexiong Luo, Jia Wu, Jian Yang, Hongyang Chen, Zhao Li, Hao Peng, and Chuan Zhou. Knowledge distillation guided interpretable brain subgraph neural networks for brain disorder exploration. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2):3559–3572, 2025.

Christopher W Lynn and Danielle S Bassett. The physics of brain network structure, function and control. *Nature Reviews Physics*, 1(5):318–332, 2019.

Hae-Jeong Park and Karl Friston. Structural and functional brain networks: From connections to cognition. *Science*, 342(6158):1238411, 2013.

Shengbing Pei, Jiajun Ma, Zhao Lv, Chao Zhang, and Jihong Guan. Community-aware graph transformer for brain disorder identification. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 4191–4199, 2025.

Ciyuan Peng, Yuelong Huang, Qichao Dong, Shuo Yu, Feng Xia, Chengqi Zhang, and Yaochu Jin. Biologically plausible brain graph transformer. In *The Thirteenth International Conference on Learning Representations*, 2025a.

Qi Peng, Yi Cai, Jiankun Liu, Quan Zou, Xing Chen, Zheng Zhong, Zefeng Wang, Jiayuan Xie, and Qing Li. Integration of multi-source medical data for medical diagnosis question answering. *IEEE Transactions on Medical Imaging*, 44(3):1373–1385, 2025b.

Michela Pievani, Nicola Filippini, Martijn P Van Den Heuvel, Stefano F Cappa, and Giovanni B Frisoni. Brain connectivity in neurodegenerative diseases—from phenotype to proteinopathy. *Nature Reviews Neurology*, 10(11):620–633, 2014.

Ting Qiu, Zhen-Qi Liu, François Rheault, Jon Haitz Legarreta, Alex Valcourt Caron, Frédéric St-Onge, Cherie Strikwerda-Brown, Amelie Metz, Mahsa Dadar, Jean-Paul Soucy, et al. Structural white matter properties and cognitive resilience to tau pathology. *Alzheimer's & Dementia*, 20(5): 3364–3377, 2024a.

Weikang Qiu, Huangrui Chu, Selena Wang, Haolan Zuo, Xiaoxiao Li, Yize Zhao, and Rex Ying. Learning high-order relationships of brain regions. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 41582–41603. PMLR, 2024b.

Gang Qu, Anton Orlichenko, Junqi Wang, Gemeng Zhang, Li Xiao, Kun Zhang, Tony W. Wilson, Julia M. Stephen, Vince D. Calhoun, and Yu-Ping Wang. Interpretable cognitive ability prediction: A comprehensive gated graph transformer framework for analyzing functional brain networks. *IEEE Transactions on Medical Imaging*, 43(4):1568–1578, 2024.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Xuegang Song, Feng Zhou, Alejandro F. Frangi, Jiuwen Cao, Xiaohua Xiao, Yi Lei, Tianfu Wang, and Baiying Lei. Multicenter and multichannel pooling GCN for early AD diagnosis based on dual-modality fused brain network. *IEEE Transactions on Medical Imaging*, 42(2):354–367, 2023.

Nicola Spotorno, Olof Strandberg, Erik Stomrud, Shorena Janelidze, Kaj Blennow, Markus Nilsson, Danielle van Westen, and Oskar Hansson. Diffusion mri tracks cortical microstructural changes during the early stages of alzheimer's disease. *Brain*, 147(3):961–969, 2024.

Haoteng Tang, Guixiang Ma, Lei Guo, Xiyao Fu, Heng Huang, and Liang Zhan. Contrastive brain network learning via hierarchical signed graph pooling model. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6):7363–7375, 2024a.

Xiangru Tang, Anni Zou, Zhuosheng Zhang, Ziming Li, Yilun Zhao, Xingyao Zhang, Arman Cohan, and Mark Gerstein. MedAgents: Large language models as collaborators for zero-shot medical reasoning. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 599–621, 2024b.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.

Qwen Team. Qwen3.5: Accelerating productivity with native multimodal agents, February 2026. URL `https://qwen.ai/blog?id=qwen3.5`.

Qianqian Wang, Wei Wang, Hong-Jun Li, Weili Lin, and Mingxia Liu. Dynamic function-structure connectivity coupling for predicting progression trajectories in neurocognitive decline. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 298–308. Springer, 2025.

Shaolong Wei, Shu Jiang, Mingliang Wang, Liang Sun, Haonan Rao, Weiping Ding, and Jiashuang Huang. SCNNs: Spike-based coupling neural networks for understanding structural-functional relationships in the human brain. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 6524–6532, 2025.

Ziquan Wei, Tingting Dan, Jiaqi Ding, and Guorong Wu. NeuroPath: A neural pathway transformer for joining the dots of human connectomes. In *Advances in Neural Information Processing Systems*, volume 37, pp. 67826–67849, 2024.

Jing Xia, Yi Hao Chan, Deepank Girish, and Jagath C Rajapakse. Interpretable modality-specific and interactive graph convolutional network on brain functional and structural connectomes. *Medical Image Analysis*, 102:103509, 2025.

Wei Xue, Hong He, Yanbing Wang, and Ying Zhao. SAGN: Sparse adaptive gated graph neural network with graph regularization for identifying dual-view brain networks. *IEEE Transactions on Neural Networks and Learning Systems*, 36(5):8085–8099, 2025.

Cheng Yang, Hui Jin, Xinlei Yu, Zhipeng Wang, Yaoqun Liu, Fenglei Fan, Dajiang Lei, Gangyong Jia, Changmiao Wang, and Ruiquan Ge. LungNoduleAgent: A collaborative multi-agent system for precision diagnosis of lung nodules. *arXiv preprint arXiv:2511.21042*, 2025a.

Meimei Yang, Yongheng Sun, Qianqian Wang, Wei Wang, Hong-Jun Li, and Mingxia Liu. Hyperbolic kernel gcn with structure-function connectivity coupling for neurocognitive impairment analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 405–415. Springer, 2025b.

Meimei Yang, Yongheng Sun, Qianqian Wang, Andrea Bozoki, Maureen Kohi, and Mingxia Liu. Hyperbolic kernel graph neural networks for neurocognitive decline analysis from multimodal brain imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(4):4827–4841, 2026.

Yanwu Yang, Chenfei Ye, Xutao Guo, Tao Wu, Yang Xiang, and Ting Ma. Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning. *IEEE Transactions on Medical Imaging*, 43(1):108–121, 2024.

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.

Ceyao Zhang, Kaijie Yang, Siyi Hu, Zihao Wang, Guanghe Li, Yihang Sun, Cheng Zhang, Zhaowei Zhang, Anji Liu, Song-Chun Zhu, et al. ProAgent: Building proactive cooperative agents with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 17591–17599, 2024.

Lu Zhang, Li Wang, Jean Gao, Shannon L. Risacher, Jingwen Yan, Gang Li, Tianming Liu, and Dajiang Zhu. Deep fusion of brain structure-function in mild cognitive impairment. *Medical Image Analysis*, 72:102082, 2021.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, et al. Qwen3 Embedding: Advancing text embedding and reranking through foundation models. *arXiv preprint arXiv:2506.05176*, 2025.

Zhixuan Zhou, Tingting Dan, and Guorong Wu. BrainFlow: A holistic pathway of dynamic neural system on manifold. In *Advances in Neural Information Processing Systems*, volume 38, pp. 32037–32067. Curran Associates, Inc., 2025.

A Data Preprocessing

dMRI Preprocessing. The dMRI data preprocessing was performed using the FSL's Diffusion Toolbox (FDT). Initially, the non-diffusion-weighted images were utilized to estimate the brain mask, effectively separating brain tissues from non-brain tissues. Subsequently, eddy-current distortion correction was applied to the diffusion-weighted images to mitigate artifacts caused by geometric

distortions and head motion during acquisition. Following these correction steps, diffusion tensors were estimated from the preprocessed images to derive the scalar metrics of fractional anisotropy and mean diffusivity. For the construction of the structural network, deterministic fiber tracking was executed using DSI Studio software. The T1-weighted anatomical images were co-registered to the diffusion space, and the Automated Anatomical Labeling (AAL) atlas was then warped into the individual diffusion space to delineate 90 regions of interest (ROIs). The structural connectivity was defined by the number of reconstructed fibers connecting each pair of ROIs. The resulting fiber-count matrices were further subjected to logarithmic transformation and global min–max normalization, followed by symmetrization. Self-connections were removed by setting all diagonal entries to zero, yielding the final SC matrix $\mathbf{A}^{\mathrm{S}} \in \mathbb{R}^{90\times 90}$.

rs-fMRI Preprocessing. The rs-fMRI data were preprocessed using the Data Processing Assistant for Resting-State fMRI (DPARSF). The first 10 time points of each participant were discarded to ensure magnetization equilibrium and subject adaptation. The remaining volumes underwent slice timing correction and realignment for head motion; participants with maximum translation exceeding 2.0 mm or rotation exceeding 2.0° were excluded from the analysis. The T1-weighted images were co-registered to the functional space and segmented into gray matter, white matter (WM), and cerebrospinal fluid (CSF). Functional images were then spatially normalized to the Montreal Neurological Institute (MNI) space using the transformation parameters estimated from T1 segmentation and resampled to $3 \times 3 \times 3$ mm^3 isotropic voxels. Subsequent processing steps included spatial smoothing with a 6 mm full-width at half-maximum Gaussian kernel, removal of linear trends, and regression of nuisance signals (including WM, CSF, and Friston-24 motion parameters). Finally, a temporal band-pass filter (0.01–0.08 Hz) was applied to reduce low-frequency drift and high-frequency noise, and the time series for each voxel were normalized to zero mean and unit variance.

For functional network construction, the preprocessed fMRI signals were parcellated according to the AAL atlas into the same 90 ROIs used for the structural network. The representative time series of each ROI was obtained by averaging the time series of all voxels within that region. Pairwise Pearson correlation coefficients were then computed between the ROI time series to construct the subject-specific functional connectivity matrix. The diagonal entries were set to zero, resulting in the final symmetric FC matrix $\mathbf{A}^{\mathrm{F}} \in \mathbb{R}^{90\times 90}$.

B Additional Implementation and RL Controller Details

This section provides additional implementation details of BrainIPAgent, including the cross-validation protocol, RL controller configuration, and training hyperparameters, to facilitate reproducibility.

Training Protocol and Data Split. BrainIPAgent is evaluated using 10-fold cross-validation. For each outer fold, the held-out test fold is used exclusively for final evaluation and is never accessed during model training, RL policy optimization, controller action selection, or checkpoint selection. Within each outer fold, 20% of the training samples are held out as an inner validation set using stratified random sampling with a fixed random seed of 72, while the remaining 80% are used for model training. The training, validation, and test subsets are mutually disjoint.

The inner training subset is used to optimize the base network, whereas the validation subset provides feedback to the RL controller and is used for model selection. The checkpoint with the highest validation accuracy is retained during 500 training epochs. After training is completed, the selected checkpoint is evaluated on the held-out test fold to obtain the final performance. Therefore, neither the RL reward nor checkpoint selection has access to the test samples, preventing information leakage from the final evaluation set.

The base network is optimized using Adam with a batch size of 100, a learning rate of 5×10^{-5}, and a weight decay of 0.1. The symmetric InfoNCE objective uses a temperature of $\tau = 0.07$, and the functional and textual representations are projected into a 128-dimensional shared embedding space. The structure-aware encoder uses nine attention heads and four hierarchical pooling stages with sizes of $[64, 32, 16, 2]$. For each subject, the interpreter agents assign importance scores to candidate patterns based on literature relevance and anatomical significance. These scores are used only for ranking and selecting the top $D = 10$ pairwise connections and top $D = 10$ triplet-based

structural patterns, rather than as continuous weights in the subsequent mask. All training and controller hyperparameters are fixed across cross-validation folds.

RL Controller State and Action Space. The RL controller operates at the epoch level. At epoch e, it observes a six-dimensional global state

$$\mathbf{s}_e = \left[\mathrm{Acc}_{e-1}^{\mathrm{val}}, \mathcal{L}_{\mathrm{CE}}^{e-1}, \mathcal{L}_{\mathrm{CL}}^{e-1}, \mathcal{H}_{\mathrm{attn}}^{e-1}, \mathcal{H}_{\mathrm{assign}}^{e-1}, \frac{e}{E-1} \right], \tag{14}$$

where $\mathrm{Acc}_{e-1}^{\mathrm{val}}$ denotes the validation accuracy from the previous epoch, $\mathcal{L}_{\mathrm{CE}}$ and $\mathcal{L}_{\mathrm{CL}}$ denote the classification and contrastive losses, $\mathcal{H}_{\mathrm{attn}}$ and $\mathcal{H}_{\mathrm{assign}}$ denote the attention and soft-assignment entropies, and $E = 500$ is the total number of training epochs. The final term represents the normalized training progress.

The policy is parameterized by a two-layer multilayer perceptron with 128 hidden units per layer and Tanh activations. It defines a Gaussian distribution over two continuous latent actions with learnable log standard deviations. The sampled actions are transformed through a sigmoid function and mapped to $\alpha_e \in [0.2, 2.0]$ and $\lambda_e \in [0.01, 0.3]$, controlling the structural guidance strength and contrastive alignment weight, respectively.

Candidate Evaluation and Reward Computation. At the beginning of each epoch, the controller samples $K = 4$ candidate actions. To ensure a fair comparison, all candidates are evaluated from the same snapshot of the base network and optimizer. For each candidate, the base network is temporarily optimized for five mini-batches from the inner training subset and subsequently evaluated on up to three validation mini-batches.

The candidate reward follows Eq. 11, with $\beta_1 = \beta_2 = 0.01$. The attention and soft-assignment entropies are averaged over the corresponding trial mini-batches. After each candidate evaluation, the base network and optimizer are restored to the common pre-trial snapshot, ensuring that candidate trials do not interfere with one another. The group-relative advantages are then computed according to Eq. 12.

Policy Optimization and Epoch-Level Training. The controller is optimized using a clipped probability-ratio objective. With $a_e^k = (\alpha_e^k, \lambda_e^k)$ denoting the k-th candidate action, the policy probability ratio is defined as

$$\rho_e^k = \frac{\pi_\theta(a_e^k \mid \mathbf{s}_e)}{\pi_{\theta_{\mathrm{old}}}(a_e^k \mid \mathbf{s}_e)}. \tag{15}$$

The policy loss is

$$\mathcal{L}_{\mathrm{policy}} = -\frac{1}{K} \sum_{k=1}^{K} \min \left[\rho_e^k A_e^k, \mathrm{clip}\left(\rho_e^k, 1-\epsilon, 1+\epsilon\right) A_e^k \right] - \eta \mathcal{H}(\pi_\theta), \tag{16}$$

where $\epsilon = 0.2$ is the clipping coefficient and $\eta = 0.01$ is the entropy regularization coefficient. The policy is optimized using Adam with a learning rate of 3×10^{-4} for four update steps per epoch, with the gradient norm clipped to 1.0.

After the policy update, the candidate with the highest reward is selected, and its (α_e, λ_e) pair is fixed throughout the subsequent full training epoch. The base network is optimized over the complete inner training subset using the objective defined in Eq. 13, with the structure-aware mask scaled by α_e. After each epoch, the full validation subset is evaluated, and the resulting validation accuracy, classification loss, contrastive loss, attention entropy, and assignment entropy are used to construct the controller state for the next epoch.

C Details of Competing Methods

We provide detailed descriptions of the eight baseline methods used in our comparative experiments. These methods are categorized based on their input modalities (single-modal vs. multimodal) and core technical methodologies.

Single-modal Methods. We include two Transformer-based methods tailored for single-modal brain network analysis, applicable to both FC and SC:

Table 3: Performance comparison with state-of-the-art methods on the ADNI dataset. We report the classification accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score (F1), and Area Under the Curve (AUC). All results are presented as Mean ± Std. The best results are highlighted in **bold**, and the second-best results are underlined.

Method	Modality	ACC (%)	SEN (%)	SPE (%)	F1 (%)	AUC (%)
BrainNetTF (Kan et al., 2022)	FC	71.54 ± 6.17	61.17 ± 13.68	80.93 ± 8.61	66.41 ± 9.33	71.40 ± 6.62
BrainNetTF (Kan et al., 2022)	SC	67.32 ± 7.93	60.60 ± 15.84	73.42 ± 12.08	62.92 ± 10.87	67.26 ± 11.64
ALTER (Yu et al., 2024)	FC	71.20 ± 7.70	61.33 ± 16.21	78.84 ± 12.83	71.06 ± 7.33	69.27 ± 13.02
ALTER (Yu et al., 2024)	SC	68.82 ± 9.80	62.52 ± 13.08	74.49 ± 14.89	68.26 ± 8.91	64.67 ± 14.99
AL-NEGAT (Chen et al., 2024)	FC, SC	71.73 ± 6.21	58.39 ± 11.35	82.49 ± 10.71	64.11 ± 8.51	70.71 ± 6.07
NeuroPath (Wei et al., 2024)	FC, SC	72.70 ± 5.47	63.17 ± 16.07	81.46 ± 11.30	67.77 ± 8.96	73.51 ± 5.89
MS-Inter-GCN (Xia et al., 2025)	FC, SC	74.62 ± 8.88	68.09 ± 11.66	80.59 ± 11.15	71.73 ± 9.29	75.82 ± 10.81
ST2GCN (Li et al., 2025)	FC, SC	74.31 ± 4.26	65.12 ± 15.14	82.73 ± 10.17	69.85 ± 7.61	76.90 ± 11.69
HKGCN (Yang et al., 2026)	FC, SC	73.86 ± 5.58	69.12 ± 14.04	78.26 ± 9.91	71.04 ± 8.12	73.87 ± 8.34
HKGAT (Yang et al., 2026)	FC, SC	76.83 ± 2.96	69.86 ± 16.45	**83.26 ± 11.91**	74.19 ± 5.34	82.94 ± 8.99
BrainIPAgent (ours)	FC, SC	**79.94 ± 5.81**	**76.83 ± 7.10**	82.83 ± 12.64	**78.60 ± 5.29**	**85.44 ± 7.10**

- **BrainNetTF** (Kan et al., 2022): To address the limitations of fixed node ordering and size in brain networks, BrainNetTF utilizes connection profiles as node features. It incorporates an Orthonormal Clustering Readout (OCREAD) mechanism to effectively extract discriminative features representing functional modules.

- **ALTER** (Yu et al., 2024): The Adaptive Long-range aware Transformer (ALTER) introduces a biased random walk strategy to explicitly capture long-range dependencies between brain regions based on correlation strengths. This approach addresses the tendency of traditional graph models to neglect distant regional synergies.

Multimodal Methods. We compare against frontier methods that fuse FC and SC. These baselines cover diverse mechanisms ranging from adversarial learning and bio-inspired pathways to spatio-temporal dynamics and hyperbolic geometry:

- **AL-NEGAT** (Chen et al., 2024): This method proposes a Node-Edge Graph Attention Network capable of simultaneously leveraging node attributes and edge connectivity strengths. Its core contribution is the integration of adversarial training strategies (FGSM and PGD), which significantly enhances model generalization and robustness on heterogeneous cross-site data.

- **NeuroPath** (Wei et al., 2024): Inspired by neuroscience, NeuroPath is a Neural Pathway Transformer. It introduces the concept of "Topological Detour," positing that FC links are supported by multi-hop SC paths. The model explicitly attends to these underlying anatomical pathways via a multi-head self-attention mechanism, revealing the complex structural support behind functional connectivity.

- **MS-Inter-GCN** (Xia et al., 2025): This method focuses on interpretable modality-specific and interactive modeling. It employs an encoder-decoder architecture to extract unique modality features and designs an interaction module to explicitly learn the coupling weights between SC and FC, constructing a unified interactive graph for prediction.

- **ST2GCN** (Li et al., 2025): Targeting dynamic brain network analysis, ST2GCN proposes a Spatio-Temporal Tensor Graph Convolutional Network. It utilizes intra-graph propagation to capture spatial topology and inter-graph propagation to model temporal evolution, incorporating functional subnetwork constraints to enhance biological interpretability.

- **HKGCN & HKGAT** (Yang et al., 2026): These methods challenge the traditional Euclidean assumption by embedding brain networks into Hyperbolic Space, which is naturally suited for representing tree-like or hierarchical structures. **HKGCN** (convolution-based) and **HKGAT** (attention-based) utilize curvature-aware kernel functions to efficiently capture the intrinsic hierarchical organization of brain networks.

Table 4: Performance comparison with state-of-the-art methods on the Zhongda dataset. We report the classification accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score (F1), and Area Under the Curve (AUC). All results are presented as Mean ± Std. The best results are highlighted in **bold**, and the second-best results are underlined.

Method	Modality	ACC (%)	SEN (%)	SPE (%)	F1 (%)	AUC (%)
BrainNetTF (Kan et al., 2022)	FC	74.87 ± 8.72	82.14 ± 11.13	65.50 ± 21.27	78.88 ± 6.56	66.65 ± 12.99
BrainNetTF (Kan et al., 2022)	SC	75.97 ± 8.13	86.67 ± 14.53	62.00 ± 14.53	79.75 ± 8.69	72.17 ± 11.79
ALTER (Yu et al., 2024)	FC	74.14 ± 5.14	86.67 ± 10.00	57.50 ± 15.37	78.23 ± 4.51	79.92 ± 9.06
ALTER (Yu et al., 2024)	SC	75.05 ± 6.91	83.33 ± 10.54	65.00 ± 20.62	78.19 ± 4.82	80.83 ± 8.86
AL-NEGAT (Chen et al., 2024)	FC, SC	78.70 ± 4.03	**90.24 ± 8.00**	64.00 ± 15.46	82.74 ± 2.77	76.76 ± 12.02
NeuroPath (Wei et al., 2024)	FC, SC	72.32 ± 8.25	81.90 ± 16.62	59.50 ± 23.27	70.82 ± 8.79	63.19 ± 12.38
MS-Inter-GCN (Xia et al., 2025)	FC, SC	77.77 ± 7.25	87.14 ± 11.77	66.00 ± 22.56	81.64 ± 5.23	73.27 ± 14.10
ST2GCN (Li et al., 2025)	FC, SC	78.59 ± 6.25	85.47 ± 8.07	69.00 ± 16.40	82.00 ± 4.64	81.94 ± 10.65
HKGCN (Yang et al., 2026)	FC, SC	75.96 ± 4.24	80.48 ± 9.45	69.50 ± 17.67	79.06 ± 3.03	77.57 ± 5.86
HKGAT (Yang et al., 2026)	FC, SC	76.87 ± 4.42	83.81 ± 12.36	67.50 ± 17.50	80.16 ± 4.10	77.40 ± 11.99
BrainIPAgent (ours)	FC, SC	**82.33 ± 5.07**	88.57 ± 7.52	**74.00 ± 13.38**	**85.06 ± 4.08**	**84.27 ± 8.58**

Table 5: Performance comparison with state-of-the-art methods on the Xinxiang dataset. We report the classification accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score (F1), and Area Under the Curve (AUC). All results are presented as Mean ± Std. The best results are highlighted in **bold**, and the second-best results are underlined.

Method	Modality	ACC (%)	SEN (%)	SPE (%)	F1 (%)	AUC (%)
BrainNetTF (Kan et al., 2022)	FC	83.63 ± 12.18	79.17 ± 23.93	86.50 ± 15.82	78.07 ± 18.31	80.50 ± 17.73
BrainNetTF (Kan et al., 2022)	SC	83.31 ± 8.08	77.50 ± 14.93	86.50 ± 15.82	78.91 ± 9.09	79.33 ± 12.72
ALTER (Yu et al., 2024)	FC	80.64 ± 8.76	74.17 ± 13.15	84.50 ± 10.36	75.41 ± 10.65	82.33 ± 12.05
ALTER (Yu et al., 2024)	SC	79.52 ± 7.31	62.50 ± 19.46	91.50 ± 10.50	69.57 ± 14.06	76.17 ± 14.32
AL-NEGAT (Chen et al., 2024)	FC, SC	88.53 ± 10.90	84.17 ± 16.01	91.00 ± 15.78	85.40 ± 13.12	85.17 ± 16.51
NeuroPath (Wei et al., 2024)	FC, SC	83.13 ± 14.40	84.17 ± 16.87	82.50 ± 15.14	83.17 ± 14.41	80.67 ± 16.60
MS-Inter-GCN (Xia et al., 2025)	FC, SC	84.74 ± 11.03	77.50 ± 14.93	89.50 ± 10.59	80.17 ± 13.91	84.33 ± 11.67
ST2GCN (Li et al., 2025)	FC, SC	88.67 ± 8.44	78.34 ± 18.33	**95.50 ± 9.07**	83.91 ± 11.98	90.00 ± 9.69
HKGCN (Yang et al., 2026)	FC, SC	84.56 ± 9.48	74.17 ± 19.88	92.00 ± 9.80	78.41 ± 14.45	80.67 ± 20.36
HKGAT (Yang et al., 2026)	FC, SC	85.99 ± 8.40	80.84 ± 15.83	89.00 ± 11.14	81.98 ± 10.98	83.50 ± 12.17
BrainIPAgent (ours)	FC, SC	**91.35 ± 7.58**	**87.50 ± 15.48**	94.00 ± 9.17	**88.64 ± 9.71**	**92.83 ± 9.25**

D Comparative Experiment Results

Evaluation Metrics. We evaluated the proposed method using Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), F1-score (F1), and Area Under the Curve (AUC).

Qwen-3.5 (Team, 2026) is used as the default LLM backbone for BrainIPAgent throughout the experiments, unless otherwise specified. **Qwen-3.5** (Team, 2026), **Gemini-3** (Team et al., 2023), and **DeepSeek-V4** (DeepSeek-AI et al., 2026) are accessed through API services, whereas the open-source **Qwen-3.5-9B** (Team, 2026) model is deployed locally.

We present the complete quantitative results for the ADNI, Zhongda, Xinxiang, and ZDXX datasets in Tables 3, 4, 5, and 6 to provide a supplementary evaluation beyond the primary accuracy metric reported in the main text. As shown in the comprehensive results, BrainIPAgent exhibits robust and consistently strong performance across all benchmarks. Specifically, our method achieves the best results in four out of five evaluation metrics on the ADNI, Zhongda, and Xinxiang datasets, and attains the top performance across all five metrics on the ZDXX dataset.

E Comparative Ablation Study Results

E.1 Complete Numerical Results for Main-Text Ablations

To complement the visual comparison in the main text, we provide the complete numerical results for the ADNI and ZDXX datasets in Table 7 and Table 8, respectively. These tables report the exact Mean ± Std values across ACC, SEN, SPE, F1, and AUC, enabling a more detailed assessment of the contribution of each component. Overall, the full BrainIPAgent consistently achieves the

Table 6: Performance comparison with state-of-the-art methods on the ZDXX dataset. We report the classification accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score (F1), and Area Under the Curve (AUC). All results are presented as Mean ± Std. The best results are highlighted in **bold**, and the second-best results are underlined.

Method	Modality	ACC (%)	SEN (%)	SPE (%)	F1 (%)	AUC (%)
BrainNetTF (Kan et al., 2022)	FC	70.82 ± 8.13	69.34 ± 14.88	72.22 ± 18.69	69.74 ± 9.92	66.55 ± 8.06
BrainNetTF (Kan et al., 2022)	SC	70.06 ± 8.30	73.00 ± 13.48	67.00 ± 18.83	70.59 ± 8.28	67.24 ± 11.11
ALTER (Yu et al., 2024)	FC	72.26 ± 7.24	74.89 ± 16.63	69.33 ± 18.69	72.40 ± 7.25	69.16 ± 11.92
ALTER (Yu et al., 2024)	SC	69.09 ± 8.79	69.78 ± 19.38	68.33 ± 16.51	69.24 ± 8.85	67.26 ± 10.72
AL-NEGAT (Chen et al., 2024)	FC, SC	77.76 ± 7.16	77.00 ± 11.35	78.67 ± 15.94	77.49 ± 6.34	75.87 ± 10.45
NeuroPath (Wei et al., 2024)	FC, SC	69.65 ± 10.57	76.11 ± 12.45	63.33 ± 16.67	69.23 ± 10.63	63.16 ± 13.34
MS-Inter-GCN (Xia et al., 2025)	FC, SC	71.18 ± 10.78	69.45 ± 18.56	72.78 ± 18.76	69.73 ± 12.71	69.55 ± 14.24
ST2GCN (Li et al., 2025)	FC, SC	77.62 ± 7.41	76.89 ± 18.25	78.34 ± 11.26	76.36 ± 9.85	73.82 ± 10.84
HKGCN (Yang et al., 2026)	FC, SC	75.42 ± 3.29	74.11 ± 12.41	76.45 ± 10.47	75.69 ± 3.66	74.33 ± 8.90
HKGAT (Yang et al., 2026)	FC, SC	75.98 ± 6.70	74.22 ± 13.14	77.67 ± 11.44	75.62 ± 7.76	76.22 ± 8.18
BrainIPAgent (ours)	FC, SC	**80.61 ± 7.16**	**77.34 ± 9.11**	**83.89 ± 9.06**	**79.84 ± 7.43**	**81.16 ± 5.61**

Table 7: Ablation study results on the ADNI dataset. We report the classification accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score (F1), and Area Under the Curve (AUC). All results are presented as Mean ± Std. The best results are highlighted in **bold**, and the second-best results are underlined.

Model	ACC (%)	SEN (%)	SPE (%)	F1 (%)	AUC (%)
Baseline (Naive Fusion)	74.95 ± 6.08	68.45 ± 12.28	80.93 ± 8.96	71.77 ± 8.14	79.41 ± 7.56
w/o Textual Alignment	76.12 ± 6.89	71.00 ± 14.70	80.89 ± 10.93	74.28 ± 7.26	79.21 ± 12.35
w/o Structural Guidance	76.95 ± 6.17	70.48 ± 10.78	81.69 ± 13.38	75.33 ± 5.24	80.35 ± 9.70
w/o RL Adaptation	77.12 ± 5.03	75.86 ± 10.78	78.34 ± 9.92	75.70 ± 5.66	81.36 ± 8.62
BrainIPAgent (ours)	**79.94 ± 5.81**	**76.83 ± 7.10**	**82.83 ± 12.64**	**78.60 ± 5.29**	**85.44 ± 7.10**

strongest performance across both datasets, showing that textual priors, structural guidance, and adaptive regulation provide complementary benefits.

Removing textual alignment leads to a consistent performance degradation and yields results close to the naive fusion baseline. This effect is particularly evident in AUC on ZDXX, where the textual-alignment variant performs even below the baseline, indicating that directly combining structural and functional information without semantic alignment is insufficient to fully exploit the agent-generated priors. Removing structural guidance also reduces overall performance, with a more noticeable decrease in SEN on both datasets. This suggests that the structure-aware mask helps preserve disease-relevant interactions and improves the sensitivity to discriminative connectivity patterns. Finally, fixing α and λ without RL adaptation still produces competitive results on several metrics, including ACC and F1, but remains consistently inferior to the full model. The improvement obtained by BrainIPAgent indicates that dynamically adjusting the strengths of structural guidance and semantic alignment better accommodates their changing roles across training stages. Together, these results further demonstrate that the three components contribute complementary information and that their joint integration is important for achieving robust performance across different evaluation criteria.

E.2 Prior Generation Strategy Ablation

Single-pass LLMs may suffer from structural hallucinations when simultaneously extracting complex graph topologies and generating clinical reasoning. The proposed multi-agent framework decouples this cognitive load by separating structural pattern extraction, clinical interpretation, and expert refinement. To explicitly isolate the benefit of this design, we conduct an additional ablation study against four prior-generation variants.

- **Single-Agent One-Shot.** This variant replaces the multi-agent framework with a single LLM that directly processes the raw structural matrix and generates the textual prior in one pass.

Table 8: Ablation study results on the ZDXX dataset. We report the classification accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score (F1), and Area Under the Curve (AUC). All results are presented as Mean ± Std. The best results are highlighted in **bold**, and the second-best results are underlined.

Model	ACC (%)	SEN (%)	SPE (%)	F1 (%)	AUC (%)
Baseline (Naive Fusion)	76.76 ± 5.45	73.78 ± 13.51	79.45 ± 14.19	76.48 ± 4.48	76.84 ± 8.23
w/o Textual Alignment	77.45 ± 5.98	75.22 ± 7.06	79.67 ± 11.47	76.81 ± 5.51	74.69 ± 6.95
w/o Structural Guidance	78.06 ± 4.38	74.11 ± 7.44	82.00 ± 9.52	77.11 ± 4.44	77.91 ± 6.22
w/o RL Adaptation	78.12 ± 4.86	76.34 ± 6.48	79.89 ± 8.78	77.64 ± 4.86	76.89 ± 7.95
BrainIPAgent (ours)	**80.61 ± 7.16**	**77.34 ± 9.11**	**83.89 ± 9.06**	**79.84 ± 7.43**	**81.16 ± 5.61**

Table 9: Ablation study on prior generation strategies on the ADNI dataset. We report the classification accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score (F1), and Area Under the Curve (AUC). All results are presented as Mean ± Std. The best results are highlighted in **bold**, and the second-best results are underlined.

Model Variant	ACC (%)	SEN (%)	SPE (%)	F1 (%)	AUC (%)
Single-Agent One-Shot	76.48 ± 5.25	71.02 ± 14.06	81.52 ± 11.42	73.54 ± 7.82	79.66 ± 7.10
Direct Graph Fusion	74.95 ± 6.08	68.45 ± 12.28	80.93 ± 8.96	71.77 ± 8.14	79.41 ± 7.56
Direct SC Thresholding	75.24 ± 5.03	74.43 ± 8.64	76.05 ± 9.44	74.14 ± 4.85	79.96 ± 7.83
Graph-Based Mask	77.32 ± 6.16	75.83 ± 8.55	78.74 ± 10.97	76.05 ± 6.37	81.85 ± 8.41
BrainIPAgent (ours)	**79.94 ± 5.81**	**76.83 ± 7.10**	**82.83 ± 12.64**	**78.60 ± 5.29**	**85.44 ± 7.10**

- **Direct Graph Fusion.** This variant directly extracts and fuses SC and FC features using Graph Transformers, without agent-based prior generation or textual prior construction. This variant corresponds to the Naive Fusion baseline in Table 2.
- **Direct SC Thresholding.** This variant replaces the Connectivity Interpreter Agents with a non-LLM magnitude-based selector. Specifically, the top 10% strongest SC connections are retained to construct the structural prior mask, while the remaining connections are masked out.
- **Graph-Based Mask.** This variant replaces the Connectivity Interpreter Agents with a GCN-based structural prior extractor. The GCN learns structural representations directly from the SC graph and generates the corresponding structural prior mask without LLM-based clinical reasoning.

As shown in Tables 9 and 10, BrainIPAgent achieves the best overall performance on both ADNI and ZDXX. Compared with Single-Agent One-Shot, the full multi-agent design brings consistent improvements, indicating that decomposing topology extraction, clinical interpretation, and expert refinement helps produce more reliable structural priors. Direct Graph Fusion performs worse than the agent-based variants, suggesting that simply fusing SC and FC features is insufficient to fully exploit the neurobiological information contained in structural connectivity. Direct SC Thresholding and Graph-Based Mask provide useful structural constraints, but lack the semantic reasoning and report-level refinement introduced by the proposed multi-agent framework. These results further support the effectiveness of multi-agent prior generation in improving both diagnostic performance and interpretability.

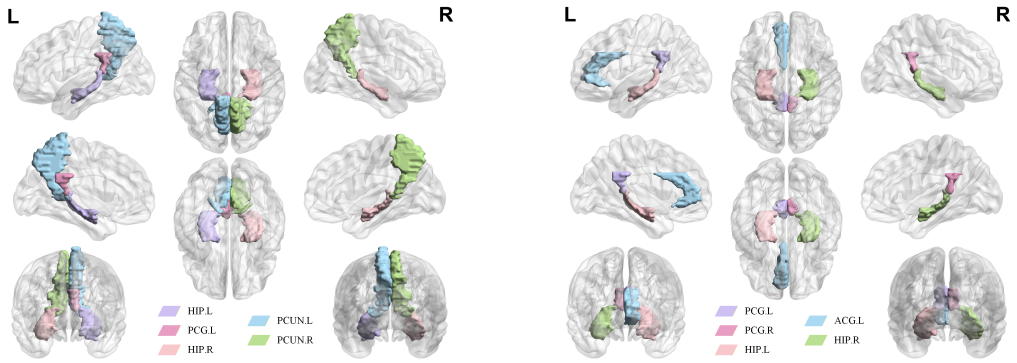
F Additional Analysis of LLM Backbones and Dialogue Dynamics

We further analyze the multi-turn refinement process across different LLM backbones from three perspectives: iterative collaboration, convergence efficiency, and robustness to model scale.

- **Impact of Iterative Collaboration.** We provide a more detailed analysis of the multi-turn refinement behavior across different LLM backbones. At $T = 0$, the generated priors are obtained without interaction with the EC Agent and therefore reflect the initial capability

Table 10: Ablation study on prior generation strategies on the ZDXX dataset. We report the classification accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-score (F1), and Area Under the Curve (AUC). All results are presented as Mean ± Std. The best results are highlighted in **bold**, and the second-best results are underlined.

Model Variant	ACC (%)	SEN (%)	SPE (%)	F1 (%)	AUC (%)
Single-Agent One-Shot	77.03 ± 5.28	71.89 ± 11.53	81.89 ± 10.66	76.51 ± 5.37	77.75 ± 8.14
Direct Graph Fusion	76.76 ± 5.45	73.78 ± 13.51	79.45 ± 14.19	76.48 ± 4.48	76.84 ± 8.23
Direct SC Thresholding	75.84 ± 3.94	74.22 ± 4.96	77.45 ± 10.28	75.44 ± 3.09	76.72 ± 7.94
Graph-Based Mask	78.01 ± 6.18	**77.45 ± 10.05**	78.56 ± 12.31	77.60 ± 6.33	80.70 ± 7.00
BrainIPAgent (ours)	**80.61 ± 7.16**	77.34 ± 9.11	**83.89 ± 9.06**	**79.84 ± 7.43**	**81.16 ± 5.61**



(a) NC vs. MCI (b) NC vs. MDD

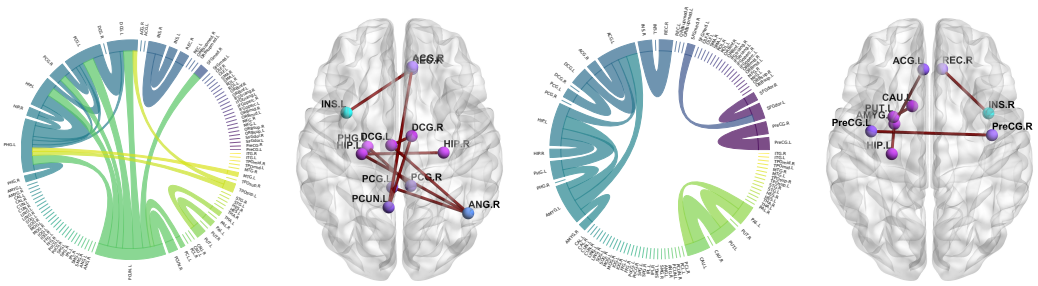
Figure 6: Visualization of the top five brain regions highlighted by PCI and HCI Agents based on Qwen-3.5. (a) NC vs. MCI classification task: HIP.L, PCG.L, HIP.R, PCUN.L, and PCUN.R. (b) NC vs. MDD classification task: PCG.L, PCG.R, HIP.L, ACG.L, and HIP.R.

of each backbone. As the dialogue proceeds, all four models show clear performance improvements, indicating that iterative expert feedback consistently benefits the quality of the generated priors. The largest gains generally occur during the first two dialogue turns, suggesting that early refinement is particularly important for correcting or improving the initial structural interpretations.

- **Convergence and Efficiency.** The performance curves further exhibit a saturation trend after $T = 2$, with later dialogue turns yielding progressively smaller improvements. This behavior indicates that the collaborative refinement process converges within only a few rounds, limiting the need for excessively long interactions. The trend is particularly relevant from an efficiency perspective, since most of the performance benefit can be obtained with a relatively small number of dialogue turns.
- **Robustness Across LLM Backbones.** We also observe consistent improvements across different LLM backbones, demonstrating that the effectiveness of BrainIPAgent is not tied to a specific foundation model. Notably, the lightweight Qwen-3.5-9B exhibits substantial gains through multi-turn refinement and gradually narrows the gap with larger models. This suggests that collaborative agent interaction can partially compensate for limited backbone capacity by progressively refining the generated structural priors and clinical interpretations.

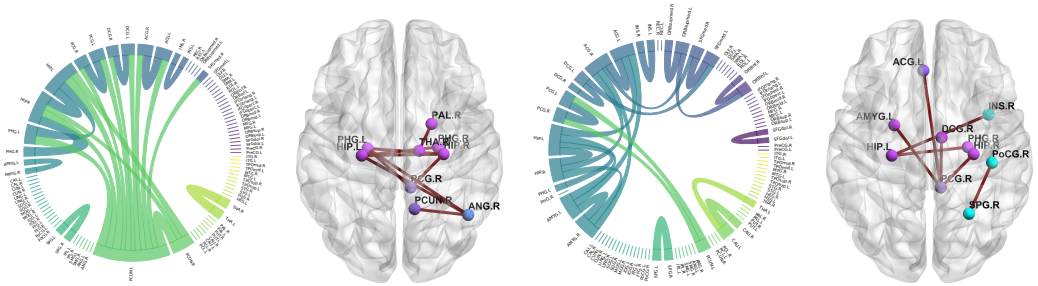
G Runtime and Training Cost Analysis

We further analyze the runtime and training cost of the computationally expensive components in BrainIPAgent. The multi-agent interpretation framework naturally introduces additional processing time compared with simpler fusion models, since it requires LLM-based structural pattern interpretation and report generation. To reduce the input length, we retain only the upper triangular part



(a) Prior (ADNI) (b) Attention (ADNI) (c) Prior (ZDXX) (d) Attention (ZDXX)

Figure 7: Visualization of interpretability results based on **Gemini-3**. (a) and (b) present the results on the ADNI (NC vs. MCI), while (c) and (d) present the results on the ZDXX (NC vs. MDD). **Prior** displays the top **structural** connections selected by the PCI and HCI Agents, and **Attention** illustrates the top structure-aware **functional** attention weights.



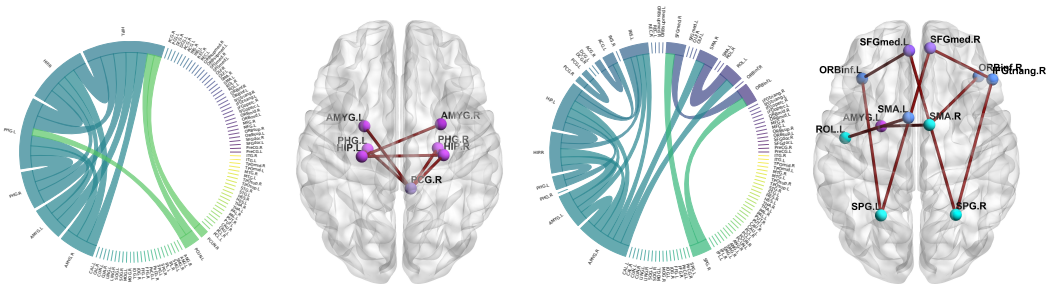
(a) Prior (ADNI) (b) Attention (ADNI) (c) Prior (ZDXX) (d) Attention (ZDXX)

Figure 8: Visualization of interpretability results based on **DeepSeek-V4**. (a) and (b) present the results on the ADNI (NC vs. MCI), while (c) and (d) present the results on the ZDXX (NC vs. MDD). **Prior** displays the top **structural** connections selected by the PCI and HCI Agents, and **Attention** illustrates the top structure-aware **functional** attention weights.

of each symmetric structural connectivity matrix, reducing the input tokens by nearly 50%. After this compression, each sample uses approximately 74K tokens on average for multi-agent prior generation.

The generation time mainly depends on the response speed of the underlying LLM service. Even under the maximum dialogue setting of $T = 5$, Gemini-3 requires an average of 34 seconds per sample, while Qwen-3.5 requires 118 seconds per sample. This indicates that the LLM overhead remains manageable even when the full multi-turn refinement process is used. This overhead is confined to the offline prior-generation stage and therefore does not affect downstream network training or inference latency once the textual priors have been precomputed. Moreover, the generated reports provide clinically meaningful structural explanations that would otherwise require manual inspection and interpretation.

After the textual priors are generated, the network training stage does not require repeated LLM calls. The RL policy updates and contrastive alignment are performed at the outer epoch level during bilevel optimization, avoiding multiplicative scaling of computational overhead. When using API-based LLMs, the training stage requires only a single NVIDIA RTX 4090 GPU. When using locally deployed LLMs such as Qwen-3.5-9B, the full pipeline can be trained with two NVIDIA RTX 4090 GPUs. On the ADNI dataset, the network training phase for 500 epochs takes only 145 seconds on average over 10 runs after the offline textual priors are obtained.



(a) Prior (ADNI) (b) Attention (ADNI) (c) Prior (ZDXX) (d) Attention (ZDXX)

Figure 9: Visualization of interpretability results based on **Qwen-3.5-9B**. (a) and (b) present the results on the ADNI (NC vs. MCI), while (c) and (d) present the results on the ZDXX (NC vs. MDD). **Prior** displays the top **structural** connections selected by the PCI and HCI Agents, and **Attention** illustrates the top structure-aware **functional** attention weights.

H Additional Interpretability Analysis

H.1 Visualization of Discriminative Brain Regions Based on Qwen-3.5.

In the NC vs. MCI task, the PCI and HCI Agents demonstrate clear disease relevance. As visualized in Figure 6(a), the agents highlight the top five discriminative regions: HIP.L, PCG.L, HIP.R, PCUN.L, and PCUN.R. These regions span medial temporal and DMN-related systems that are frequently implicated in cognitive decline and altered brain connectivity in MCI and Alzheimer's disease (Pievani et al., 2014; Lo et al., 2010).

Similarly, in the NC vs. MDD task, the agents identify PCG.L, PCG.R, HIP.L, ACG.L, and HIP.R as the top five discriminative regions, as shown in Figure 6(b). These regions involve cingulate and medial temporal systems associated with affective regulation and self-referential processing, which have been widely implicated in altered connectivity in depression (Drysdale et al., 2017).

H.2 Topological Interpretability Results on Additional Backbones.

To validate the generalization ability and robustness of our method, we extend the interpretability analysis to three additional LLM backbones: Gemini-3, DeepSeek-V4, and Qwen-3.5-9B. Across these architectures, the PCI and HCI Agents demonstrate robust consistency in identifying core pathological circuits while exhibiting architecture-specific sensitivity to distinct disease features. In the NC vs. MCI task, the agents universally anchor on the DMN and medial temporal lobe hubs, adapting to uncover nuances ranging from subcortical relay dysfunction to memory-emotion circuit integration. In the NC vs. MDD task, the agents reveal a multi-dimensional spectrum of the Cortico-Limbic circuit, elucidating pathological heterogeneities that range from striatal-motor patterns to long-range top-down regulatory deficits. Specific interpretability results for these models are detailed below:

- **Gemini-3.** In the NC vs. MCI task, the PCI and HCI Agents prioritize structural connections among DMN-related and medial temporal regions, including HIP, PHG, PCG, and PCUN in Figure 7(a). Guided by this agent-selected topology, the structure-aware attention captures the functional synchronization within these posterior regions in Figure 7(b), highlighting the coupling between the bilateral hippocampus and the posterior cingulate cortex. For the NC vs. MDD task, the agents select a distinct set of links within the Cortico-Limbic-Striatal circuit, encompassing regulatory nodes like ACG and REC, emotional regions like INS, and motor-related areas including PreCG and CAU in Figure 7(c). Subsequently, the structure-aware attention identifies specific functional coupling patterns along this pathway in Figure 7(d), revealing the interaction between the frontal-limbic system and the sensorimotor network.
- **DeepSeek-V4.** In the NC vs. MCI task, the PCI and HCI Agents prioritize structural connections linking DMN-related and medial temporal regions with subcortical nuclei,

system_content:
You are a clinical neuroimaging specialist with expertise in structural connectivity alterations associated with mild cognitive impairment (MCI).
Task Background: You are given the structural connectivity matrix of a subject whose clinical status is unknown (either MCI or healthy control). The data consist of 90 rows corresponding to the 90 AAL regions. To reduce context length, only the upper triangular matrix (including the diagonal, already set to 0) is provided. The lower triangular entries are symmetric, and row i contains values for columns $j \geq i$ only.
Your Task: Identify the most diagnostically informative connections in the matrix based on known structural connectivity alterations reported in MCI research.
Important Notes:
1. Since no population-level normative data are provided, do not infer "abnormality" directly from the magnitude of the weights.
2. Prioritize connections known to be relevant to MCI pathophysiology, but avoid selecting too many connections from the same region.
3. Zero-valued connections must not be ignored; if a key anatomical pathway shows zero connectivity, this may suggest early network-level disruption relevant to cognitive decline.
Output Requirements: For each selected connection, assign an importance score (0-1). This score should be determined by a combination of literature relevance (how frequently the connection is discussed in MCI research) and anatomical importance, and it must not rely solely on the magnitude of the matrix values.
Output Format: Return a JSON list following the structure shown below, with no additional explanation:
```
[
  {"i": 58, "j": 59, "importance": 0.85},
  {"i": 25, "j": 28, "importance": 0.72}
]
```
Use only the format and do not copy any example content.

user_content:
Based on the following data, identify the top {top_k} connections with the greatest potential diagnostic value.
Input Data: [Structural connectivity matrix]

Figure 10: Prompt template for the Pairwise Connectivity Interpreter Agent.

specifically selecting links involving HIP, PHG, PCG, and PCUN, as well as THA in Figure 8(a). Guided by this topology, the structure-aware attention captures functional synchronization in Figure 8(b), highlighting the critical coupling between cortical memory centers and subcortical relay stations. For the NC vs. MDD task, the agents select a distinct set of links spanning limbic, cingulate, and sensory-parietal networks, involving ACG, AMYG, HIP, and INS, alongside DCG, and SPG in Figure 8(c). Subsequently, the structure-aware attention identifies specific functional coupling patterns along this pathway in Figure 8(d), revealing the significant interaction between the emotion regulation circuit and the somatosensory-parietal system.

- **Qwen-3.5-9B.** In the NC vs. MCI task, the PCI and HCI Agents concentrate on the medial temporal lobe and its posterior connections, actively selecting structural links between HIP, PHG, and AMYG in Figure 9(a). Upon this topological foundation, the structure-aware attention map in Figure 9(b) reveals a dense functional interplay, particularly emphasizing the synchronization between memory circuits and the amygdala, suggesting an integrated biomarker of memory and emotional processing. For the NC vs. MDD task, the agents isolate a widely distributed network connecting frontal executive regions like SFGmed and ORBinf with sensorimotor and parietal areas such as SMA, ROL, and SPG in Figure 9(c). The corresponding functional weighting in Figure 9(d) illuminates long-range coupling patterns, effectively capturing the disruption in top-down regulation from the prefrontal cortex to posterior parietal and limbic systems.

I Prompt Templates for the Multi-Agent Framework

To ensure the reproducibility of our work and provide transparency regarding the behavior of the agents, we present the detailed prompt templates used in our framework. Figures 10, 11, 12, 13, and 14 illustrate the specific instructions, role definitions, and task constraints provided to each agent.

The prompts shown here correspond to the NC vs. MCI setting. For the NC vs. MDD experiments on the Zhongda, Xinxiang, and ZDXX datasets, the same prompt structure, agent roles, output format, and refinement procedure are used, with the disease-specific clinical context adapted from MCI to MDD. The MDD-specific prompts are provided in our public code repository.

J Broader Impact

This work aims to improve interpretable multimodal brain network learning for brain disorder analysis. Its potential positive impact lies in supporting transparent neuroimaging research and helping identify clinically meaningful structural-functional connectivity patterns. Potential risks include over-reliance on automated predictions, site-specific bias, privacy concerns associated with neuroimaging data, and possible misinterpretation of LLM-generated reports. BrainIPAgent is intended as a research tool rather than a standalone clinical decision system, and its outputs should be interpreted with clinical expertise and validated through further prospective studies.

system _content:
You are a clinical neuroimaging specialist with expertise in structural connectivity (SC) alterations associated with mild cognitive impairment (MCI).
Task Background: You are given the full structural connectivity matrix of a subject whose clinical status is unknown (either MCI or healthy control). The data consist of 90 rows corresponding to the 90 AAL regions. To reduce context length, only the upper triangular matrix (including the diagonal, already set to 0) is provided. The lower triangular entries are symmetric, and row i contains values for columns j ≥ i only.
Your Task: Based on network-level pathophysiological mechanisms of MCI reported in the neuroimaging literature, identify the most diagnostically informative higher-order triplets (i < j < k) from this matrix.
Important Notes:
1. Since no population-level normative data are provided, do not infer "abnormality" directly from the magnitude of the weights.
2. Focus on anatomical triplets that play a critical role in MCI pathophysiology, but avoid selecting too many triplets that repeatedly involve the same region.
3. Zero-valued connections must not be ignored; if a key anatomical pathway shows zero connectivity, this may suggest early network-level disruption relevant to cognitive decline.
4. Output only unique triplets satisfying i < j < k. Do not output different permutations of the same triplet and do not include any repeated nodes within a triplet.
Output Requirements: For each selected connection, assign an importance score (0-1). This score should be determined by a combination of literature relevance (how frequently the connection is discussed in MCI research) and anatomical importance, and it must not rely solely on the magnitude of the matrix values.
Output Format: Return a JSON list following the structure shown below, with no additional explanation:
```
[
  {"i": 14, "j": 23, "k": 59, "importance": 0.85},
  {"i": 16, "j": 19, "k": 41, "importance": 0.72}
]
```
Use only the format and do not copy any example content.

user_content:
Based on the following data, identify the top {top_k} high-order triplets (i < j < k) with the greatest potential diagnostic value.
Input Data: [Structural connectivity matrix]

Figure 11: Prompt template for the High-Order Connectivity Interpreter Agent.

Initial Generation
system _content:
You are a clinical neuroimaging specialist with expertise in structural connectivity (SC) alterations associated with mild cognitive impairment (MCI).
Task Background: You are provided with a list of selected pairwise structural connections from a subject whose diagnostic status is unknown (either MCI or healthy control). Each entry includes the specific brain regions involved and the corresponding connection strength (sc_strength).
Your Task: Based on the provided pairwise structural connections, write a concise clinical interpretation focusing on the meaning of the connection strength (sc_strength) and its potential neurobiological implications.
Important Notes:
1. Since no population-level normative data are available, do not assume that the subject is a patient.
2. Clearly identify the involved brain regions and the networks they may belong to.
3. If a connection known to be important in the literature appears with very low strength, still give attention to it, as this may suggest potential pathological disconnection.
Output Requirements: Do not output your reasoning. Provide a single, coherent paragraph written in appropriate neuroscientific language. The summary must be within 200 words, avoid bullet points, and avoid redundancy.

user_content:
Input Data: [Pairwise connections]

Refinement
system _content:
You are a clinical neuroimaging specialist with expertise in structural connectivity (SC) alterations associated with mild cognitive impairment (MCI).
Your Task: Revise the previous clinical interpretation report on pairwise connections based on feedback from the senior attending physician.
Important Notes:
1. You must incorporate the reviewer's key recommendations, but do not explicitly address or respond to the reviewer within the report. The revisions should be reflected naturally in the updated text.
2. Remove redundant or repetitive expressions to ensure the report is concise, clear, and aligned with clinical writing standards.
Input Description:
1. Previous report: the text that needs to be revised.
2. Reviewer's suggestions: the essential points that must be reflected in the updated report.
3. Pairwise connections: to ensure consistency and accuracy in anatomical terminology.
Output Requirements: Do not output your reasoning. Provide a single, coherent paragraph written in appropriate neuroscientific language. The summary must be within 200 words, avoid bullet points, and avoid redundancy.

user_content:
Input Data: [Previous report] [Reviewer suggestions] [Pairwise connections]

Figure 12: Prompt template for the Pairwise Connectivity Clinician Agent.

Initial Generation
system _content:
You are a clinical neuroimaging specialist with expertise in structural connectivity (SC) alterations associated with mild cognitive impairment (MCI).
Task Background: You are provided with a list of key high-order triplet structures from a subject whose diagnostic status is unknown (either MCI or healthy control). Each triplet consists of three brain regions and includes the strengths (sc_strength) of the three corresponding edge-level structural connections.
Your Task: Based on the provided triplet structures, write a concise clinical interpretation focusing on the connection strengths (sc_strength) of the three edges and their potential clinical significance.
Important Notes:
1. Since no population-level normative data are available, do not assume that the subject is a patient.
2. Clearly identify the involved brain regions and the networks they may belong to.
3. If a connection known to be important in the literature appears with very low strength, still give attention to it, as this may suggest potential pathological disconnection.
Output Requirements: Do not output your reasoning. Provide a single, coherent paragraph written in appropriate neuroscientific language. The summary must be within 200 words, avoid bullet points, and avoid redundancy.

user_content:
Input Data: [High-order triplets]

Refinement
system _content:
You are a clinical neuroimaging specialist with expertise in structural connectivity (SC) alterations associated with mild cognitive impairment (MCI).
Your Task: Revise the previous clinical interpretation report on high-order triplet structural relationships based on feedback from the senior attending physician.
Important Notes:
1. You must incorporate the reviewer's key recommendations, but do not explicitly address or respond to the reviewer within the report. The revisions should be reflected naturally in the updated text.
2. Remove redundant or repetitive expressions to ensure the report is concise, clear, and aligned with clinical writing standards.
Input Description:
1. Previous report: the text that needs to be revised.
2. Reviewer's suggestions: the essential points that must be reflected in the updated report.
3. High-order triplets: to ensure consistency and accuracy in anatomical terminology.
Output Requirements: Do not output your reasoning. Provide a single, coherent paragraph written in appropriate neuroscientific language. The summary must be within 200 words, avoid bullet points, and avoid redundancy.

user_content:
Input Data: [Previous report] [Reviewer suggestions] [High-order triplets]

Figure 13: Prompt template for the High-Order Connectivity Clinician Agent.

system _content:
You are a clinical neuroimaging specialist with expertise in structural connectivity (SC) alterations associated with mild cognitive impairment (MCI).
Task Background: You are responsible for reviewing two clinical interpretive reports written by junior physicians.
Your Task: Evaluate both the pairwise and high-order reports in terms of professional accuracy (consistency with neuroscientific and structural connectivity principles) and readability (clarity, organization, and adherence to clinical writing norms). Provide a score and revision suggestions for each report.
Review Targets: Pairwise report: describes structural connectivity between two brain regions. High-order report: describes a local circuit formed by three regions and their associated connectivity features.
Important Notes:
1. The dataset includes both patients and healthy controls. The current subject may belong to either group. Do not assume the subject is a patient. When scoring, focus on whether the report accurately reflects network characteristics relevant to cognitive and memory-related systems.
2. All comments and suggestions must strictly adhere to the provided original connection lists (Pairwise connections / High-order triplet list). Do not invent regions or connections that do not exist.
3. Evaluate readability and identify any unclear statements, factual errors, logical inconsistencies, or expressions that violate clinical writing conventions.
4. If the current report fails to implement revisions suggested in previous rounds (if applicable), you must point out omissions or incorrect modifications.
Output Requirements:
1. Provide a score from 0-1 for the pairwise report, rounded to two decimal places.
2. Provide a score from 0-1 for the high-order report, rounded to two decimal places.
3. Provide concise revision suggestions for both the pairwise and high-order reports, each consisting of 1-2 sentences.
4. Score strictly based on the current reports' quality. Do not mimic or reproduce scores from past examples.
5. Your suggestions must correspond directly to the specific content of each report and must not use generic or template-like statements.
Output Format:
PAIRWISE_SCORE: (score for pairwise report)
PAIRWISE_SUGGESTIONS: (suggestions for pairwise report)
HIGH_ORDER_TRIPLET_SCORE: (score for high-order report)
HIGH_ORDER_SUGGESTIONS: (suggestions for high-order report)
Do not output your reasoning.

user_content:
This is review round {round_id}.
Input Data: Previous Review Record (if applicable): [Previous pairwise score] [Previous pairwise suggestions] [Previous high-order score] [Previous high-order suggestions]
Content to be reviewed in the current round: [Pairwise report] [High-order report]
For reference, below are the original connection lists used to generate these reports: [Pairwise connections] [High-order triplets]

Figure 14: Prompt template for the Expert Commentator Agent.