BrainSTR: Spatio-Temporal Contrastive Learning for Interpretable Dynamic Brain Network Modeling

Guiliang Guo[1], Guangqi Wen[2], Lingwen Liu[3], Ruoxian Song[1], Peng Cao[1(✉)], Jinzhu Yang[1], Fei Wang[4(✉)], Xiaoli Liu[5], and Osmar R. Zaiane[6]

[1] Computer Science and Engineering, Northeastern University, Shenyang, China
caopeng@cse.neu.edu.cn
[2] School of Computer Science and Artificial Intelligence, Shandong Normal University, Jinan, China
[3] Graduate School of Information, Production and Systems, Waseda University, Kitakyushu, Fukuoka, Japan
[4] Nanjing Medical University, Nanjing, China
[5] AiShiWeiLai AI Research, Beijing, China
[6] Amii, University of Alberta, Edmonton, Alberta, Canada

Abstract. Dynamic functional connectivity captures time-varying brain states for better neuropsychiatric diagnosis and spatio-temporal interpretability, i.e., identifying *when* discriminative disease signatures emerge and *where* they reside in the connectivity topology. Reliable interpretability faces major challenges: diagnostic signals are often subtle and sparsely distributed across both time and topology, while nuisance fluctuations and non-diagnostic connectivities are pervasive. To address these issues, we propose **BrainSTR**, a spatio-temporal contrastive learning framework for interpretable dynamic brain network modeling. BrainSTR learns state-consistent phase boundaries via a data-driven *Adaptive Phase Partition* module, identifies diagnostically critical phases with attention, and extracts disease-related connectivity within each phase using an *Incremental Graph Structure Generator* regularized by binarization, temporal smoothness, and sparsity. Then, we introduce a spatio-temporal supervised contrastive learning approach that leverages diagnosis-relevant spatio-temporal patterns to refine inter-sample similarity and capture discriminative spatio-temporal features, constructing a well-structured semantic space for interpretable representations. Experiments on ASD, BD, and MDD validate the effectiveness of BrainSTR, and the discovered critical phases and subnetworks provide interpretable evidence. Our code: https://github.com/guoguiliang111/BrainSTR.

Keywords: Spatio-temporal interpretability · Brain-state phase partitioning · Topological filtering · Contrastive learning.

1 Introduction

Functional brain network analysis using resting-state fMRI (rs-fMRI) has shown great potential for assisting the diagnosis of neuropsychiatric disorders such as

ASD, BD, and MDD [17, 20]. To characterize time-varying brain states, dynamic functional connectivity (dFC) is often estimated by constructing a sequence of time-resolved networks from BOLD signals, commonly using fixed-length sliding-window schemes or adaptive temporal partitioning strategies [11, 15, 5]. Despite its ability to model dynamic brain states, a key challenge in applying dFC to neuropsychiatric diagnosis is achieving spatio-temporal interpretability, i.e., pinpointing precisely when and where in the connectivity topology discriminative disease signatures emerge, to uncover reliable biomarkers. Reliable spatio-temporal interpretability from dynamic brain networks remains challenging because disease-related signals are often subtle and sparsely distributed across both time and topology, whereas disease-irrelevant fluctuations and connectivities are pervasive.

To address these issues, we propose BrainSTR, a spatio-temporal contrastive learning framework for interpretable dynamic brain network modeling. Conventional contrastive learning suffers from interference by excessive disease-irrelevant information in brain networks, which compromises similarity estimation and leads to reduced similarity even among semantically similar subjects (positive pairs). In contrast, BrainSTR constructs a well-structured semantic space by jointly selecting temporally critical phases and identifying topologically disease-relevant connectivity patterns. Specifically, we learn state-consistent phase boundaries via a data-driven Adaptive Phase Partition (APP) module and identify a compact set of diagnostically critical phases with attention. Meanwhile, we propose an incremental graph structure generator, which is regularized with binarization, temporal smoothness, and sparsity terms to yield stable and interpretable connectivity selection for retaining disease-related connectivities while filtering out disease-unrelated topology within each phase. With the spatio-temporal disease-related representations, we apply supervised contrastive learning to encourage an optimal spatio-temporal semantic embedding space where the intrinsic similarity patterns of the spatio-temporal data are captured. Finally, this overall spatio-temporal interpretable representation learning is jointly optimized through critical-phase identification and incremental graph structure generator to construct a disease-relevant embedding space. Experiments on ASD, BD, and MDD show that BrainSTR achieves the best ACC/AUC on all three tasks, reaching 77.2%/77.8% for MDD, 78.2%/79.6% for BD, and 72.4%/73.0% for ASD (NYU). Compared with the strongest prior results, BrainSTR improves performance by +2.9%/+6.5% on MDD, +4.5%/+4.9% on BD, and +3.0%/+3.3% on ASD (ACC/AUC), respectively. These consistent gains indicate robust cross-disorder generalization with clear performance margins. Moreover, the discovered critical phases and subnetworks provide interpretable evidence consistent with prior neuroimaging findings. **Our main contributions are:**

1. We propose **BrainSTR**, a contrastive learning framework for dynamic brain network diagnosis that shifts the focus from redundant whole-sample representations to the sparse yet critical disease-relevant spatio-temporal features.

2. We introduce a critical **spatio-temporal learning** scheme, involving adaptive phase partition for capturing diagnostically informative phases and incremental graph structure generator for learning disease-relevant connectivity patterns, reducing temporal noise and topological redundancy to improve robustness and interpretability.
3. Extensive experiments on ASD, BD, and MDD demonstrate that BrainSTR **consistently outperforms state-of-the-art methods**, while ablation studies verify the necessity of each component. Moreover, BrainSTR discovers **clinically meaningful critical phases** and brain subnetworks that are consistent with prior neuroimaging findings, providing strong spatio-temporal interpretability.

2 Method

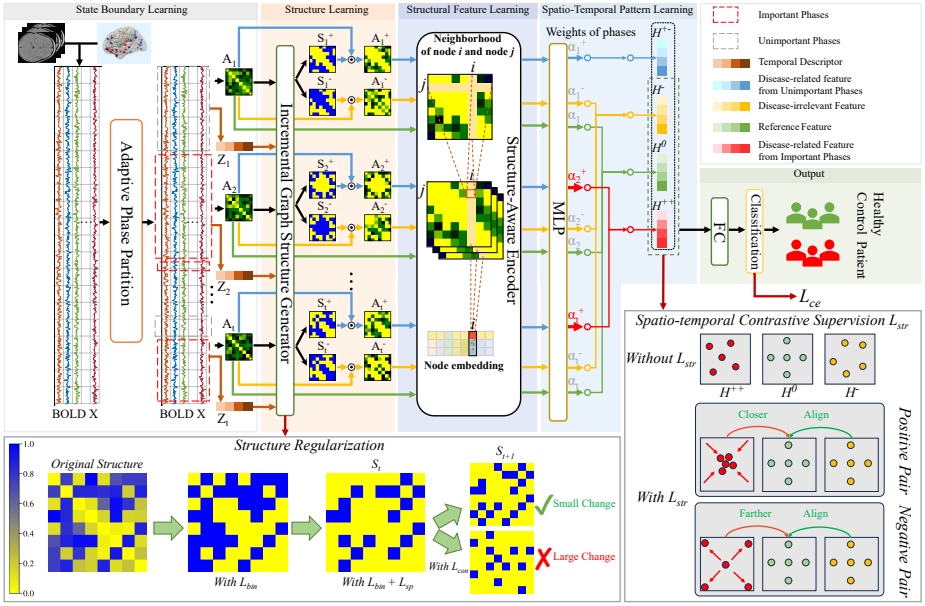


Fig. 1. Overview of the proposed **BrainSTR** framework. Given a subject's BOLD signal $X \in \mathbb{R}^{T\times N}$, where T and N denote the numbers of time points and ROIs, respectively, Adaptive Phase Partition (APP) infers *state-consistent* phase boundaries to construct phase-wise FCs $\{A_t\}_{t=1}^{W}$, where W is the number of partitioned phases, with each $A_t \in \mathbb{R}^{N\times N}$. Each A_t is decomposed into A_t^+/A_t^- via an Incremental Graph Structure Generator that learns a phase-wise structure S_t, and then encoded by a shared Structure-Aware Encoder. Attention then weights informative phases to form subject embeddings, while contrastive supervision makes them diagnosis-discriminative relative to original-graph embeddings.

2.1 Adaptive Phase Partition via Time-Invariant Representation

As shown in Fig. 1, we construct an overlapping segment sequence by sliding windowing (size w, step s=1), with w=200 for MDD/BD and w=40 for ASD, yielding $K = \lfloor \frac{T-w}{s} \rfloor + 1$ segments $\{X_k\}_{k=1}^{K}$, $X_k \in \mathbb{R}^{w \times N}$. To perform boundary learning, we train a temporal autoencoder to encode each segment as $\mathbf{h}_k = f_{\text{enc}}(X_k) \in \mathbb{R}^d$ and disentangle $\mathbf{h}_k = [\mathbf{s}_k, \mathbf{u}_k]$ into time-invariant $\mathbf{s}_k$ and time-variant $\mathbf{u}_k$ components. We implement f_{enc} as a dilated TCN and use a symmetric decoder to reconstruct $\hat{X}_k = f_{\text{dec}}(\mathbf{h}_k)$. The training objective is:

$$\mathcal{L}_{\text{APP}} = \mathcal{L}_{\text{recon}} + \lambda_{\text{smooth}}\mathcal{L}_{\text{smooth}} + \lambda_{\text{orth}}\mathcal{L}_{\text{orth}}, \tag{1}$$

where $\mathcal{L}_{\text{recon}}$ enforces reconstruction of BOLD segments, $\mathcal{L}_{\text{smooth}}$ penalizes abrupt changes of $\mathbf{s}_k$ across segments to encourage stable brain-state trajectories, and $\mathcal{L}_{\text{orth}}$ encourages disentanglement between $\mathbf{s}_k$ and $\mathbf{u}_k$ to reduce state–noise leakage. Changepoints are detected on $\{\mathbf{s}_k\}$ when $d_k > \tau_c$ (with $d_k = \|\mathbf{s}_k - \mathbf{s}_{k-1}\|_2^2$), yielding W phases with boundaries on the original BOLD time axis $0 = c_0 < \cdots < c_W = T$, where τ_c is a threshold and $\{c_t\}_{t=0}^{W}$ are the detected boundary indices. Unlike standard change-point detection that operates on raw noisy BOLD signals and tends to over-segment, APP detects changepoints in the denoised time-invariant subspace $\{\mathbf{s}_k\}$, yielding robust state boundaries. Each adjacent changepoint pair (c_{t-1}, c_t) defines one phase, and we compute the phase-wise FC sequence $\{A_t\}_{t=1}^{W}$ by Pearson correlation within each phase for subsequent modules. We set $\lambda_{\text{smooth}} = 0.1$ and $\lambda_{\text{orth}} = 1.0$ for all tasks, and select τ_c from $\{0.01, 0.05, 0.1, 0.5\}$ on the validation set (per dataset).

2.2 Incremental Graph Structure Generator

For each $A_t \in \mathbb{R}^{N \times N}$ $(t = 1, \ldots, W)$, we learn a graph structure S_t to separate disease-relevant from irrelevant connectivities.

Base and incremental structure. We learn a trainable base structure $S_0 \in \mathbb{R}^{N \times N}$, initialized to a small positive constant to avoid biasing the model toward any specific connectivity. For each phase, an MLP f_θ predicts an incremental update $\Delta S_t \in \mathbb{R}^{N \times N}$ determined by a compact temporal descriptor $\mathbf{z}_t \in \mathbb{R}^3$, defined as $\mathbf{z}_t = \left[\frac{c_{t-1}+c_t}{2T}, \frac{c_t - c_{t-1}}{T}, \|A_t - A_{t-1}\|_F\right]$, where the three terms encode normalized location, duration, and FC change magnitude. Given $\mathbf{z}_t$, the structure evolves recursively as $\Delta S_t = f_\theta(\mathbf{z}_t)$ and $S_t = S_{t-1} + \alpha_\Delta \Delta S_t$. This incremental design captures progressive topological changes across phases with shared parameters and naturally supports variable-length sequences. We set $\alpha_\Delta = 0.01$ and choose the initialization constant (0.05) on the validation set.

Differentiable structure learning. We adopt a straight-through estimator (STE) to separate disease-related from disease-irrelevant connectivities via end-to-end trainable manner: $\tilde{S}_t = \mathbb{1}(S_t > 0.5)$ in the forward pass, with gradients backpropagated through S_t. We then obtain $A_t^+ = \tilde{S}_t \odot A_t$ and $A_t^- = (\mathbf{1} - \tilde{S}_t) \odot A_t$,

where $\odot$ denotes the element-wise (Hadamard) product; A_t^+ retains selected disease-related connectivities and A_t^- captures the complementary topological patterns.

Structure regularization. We regularize the structures with three terms: a binarization loss $\mathcal{L}_{\text{bin}} = \sum_{t=1}^{W} \|S_t \odot (1-S_t)\|_F^2$ that encourages near-binary structures, a temporal consistency loss $\mathcal{L}_{\text{ms}} = \sum_{t=2}^{W} \log\big(1+\exp(\|S_t - S_{t-1}\|_F^2 - \delta)\big)$ that promotes smooth evolution across phases, where the margin δ allows mild variations but penalizes abrupt changes, and a sparsity loss $\mathcal{L}_{\text{sp}} = \sum_{t=1}^{W} \frac{1}{N^2} \|S_t\|_1$, which encourages sparse disease-related topology and is consistent with evidence that large number of connectivities are not statistically associated with disease [24]. We set $\delta = 0.1$ on the validation set.

2.3 Structure-Aware Encoder

For each $A_t^{(\cdot)} \in \{A_t^+, A_t^-, A_t\}$, we compute a phase embedding using a shared Structure-Aware Encoder $g(\cdot)$. We first apply an edge-to-edge operator $E_t^{(\cdot)} = \text{Conv}_{N,1}(A_t^{(\cdot)}) + \text{Conv}_{1,N}(A_t^{(\cdot)})$ to capture structured interactions among ROI pairs, and then use an edge-to-node operator $F_t^{(\cdot)} = \text{Conv}_{1,N}(E_t^{(\cdot)})$ to aggregate edge features. Finally, we obtain a topology-aware embedding $\mathbf{h}_t^{(\cdot)} = g(A_t^{(\cdot)}) = \text{MLP}(\text{Flatten}(F_t^{(\cdot)})) \in \mathbb{R}^d$, where d denotes the embedding dimension. The encoder is shared across $\{A_t^+, A_t^-, A_t\}$ to enforce consistent representations and reduce parameters, producing $\mathbf{h}_t^+, \mathbf{h}_t^-$, and $\mathbf{h}_t^0$, respectively.

2.4 Spatio-Temporal Contrastive Learning

Attention and aggregation. Given phase embeddings $\{\mathbf{h}_t^+, \mathbf{h}_t^-, \mathbf{h}_t^0\}_{t=1}^W$ from the shared subgraph encoder, we assign each phase an importance weight with $\alpha_t^{(\cdot)} = \frac{\exp\left(f_{\text{attn}}(\mathbf{h}_t^{(\cdot)})\right)}{\sum_{k=1}^{W} \exp\left(f_{\text{attn}}(\mathbf{h}_k^{(\cdot)})\right)}$, $(\cdot) \in \{+, -, 0\}$, where f_{attn} is implemented as a two-layer MLP with a Tanh nonlinearity that maps an embedding to a scalar score. We identify important phases by $\mathcal{I} = \{t \mid \alpha_t^+ > 1/W\}$ (or $\{\arg\max_t \alpha_t^+\}$ if empty). The uniform baseline $1/W$ represents equal phase contribution; exceeding it provides a principled, parameter-free criterion rather than a heuristic threshold. We normalize $\tilde{\alpha}_t^+ = \alpha_t^+ / \sum_{k \in \mathcal{I}} \alpha_k^+$. We then aggregate disease-related connectivity from important phases as $\mathbf{H}^{++} = \sum_{t \in \mathcal{I}} \tilde{\alpha}_t^+ \mathbf{h}_t^+$. Meanwhile, the original graph embedding $\mathbf{H}^0 = \sum_{t=1}^{W} \alpha_t^0 \mathbf{h}_t^0$ is obtained for contrastive supervision. For the complementary representation, we define $\mathbf{H}^- = \sum_{t=1}^{W} \alpha_t^- \mathbf{h}_t^-$, which is later aligned with $\mathbf{H}^0$ via a reference-guided contrastive term. The classifier predicts $\hat{\mathbf{y}} = \text{softmax}(W_c \mathbf{H}^{++} + \mathbf{b}_c)$, where W_c and $\mathbf{b}_c$ are the learnable weight matrix and bias of the final linear classifier.

Spatio-temporal contrastive loss. With the original graph embedding $\mathbf{H}^0$ as reference, we define an InfoNCE-style objective with two terms, $\mathcal{L}_{\text{ref}}$ and $\mathcal{L}_{\text{usl}}$:

$$\mathcal{L}_{\text{str}} = w_{\text{ref}} \mathcal{L}_{\text{ref}} + w_{\text{usl}} \mathcal{L}_{\text{usl}}, \tag{2}$$

$$\mathcal{L}_{\text{ref}} = -\sum_i \frac{1}{|\mathcal{P}(i)|} \sum_{p\in\mathcal{P}(i)} \log \frac{\exp(\ell_{ip}^{\text{ref}})}{\sum_{j\neq i}\exp(\ell_{ij}^{\text{ref}})}, \tag{3}$$

$$\mathcal{L}_{\text{usl}} = -\sum_i \log \frac{\exp(\ell_{ii}^{\text{usl}})}{\sum_j \exp(\ell_{ij}^{\text{usl}})}. \tag{4}$$

where $\mathcal{L}_{\text{ref}}$ forces the model to focus exclusively on disease-specific deviations by removing the inherent similarity estimated by $\mathbf{H}^0$ (representing non-diagnostic, healthy cohort-shared background dynamics), and $\mathcal{L}_{\text{usl}}$ aligns the disease-irrelevant spatio-temporal representation $\mathbf{H}^-$ with $\mathbf{H}^0$, so $\mathbf{H}^-$ remains stable and forces $\mathbf{H}^{++}$ to faster convergence during training. This is critical because $\mathbf{H}^{++}$ and $\mathbf{H}^-$ are jointly learned from complementary structures (A_t^+ and A_t^-). Together, $\mathcal{L}_{\text{ref}}$ and $\mathcal{L}_{\text{usl}}$ encourage $\mathbf{H}^{++}$ to focus on diagnosis-relevant spatio-temporal patterns. We use temperature-scaled similarity scores, with $\ell_{ij}^{\text{ref}} = \big(\text{sim}(\mathbf{H}_i^{++}, \mathbf{H}_j^{++}) - \beta\,\text{sim}(\mathbf{H}_i^0, \mathbf{H}_j^0)\big)/\tau$. Similarly, $\ell_{ij}^{\text{usl}} = \text{sim}(\mathbf{H}_i^0, \mathbf{H}_j^-)/\tau$. Here $\mathcal{P}(i) = \{p \neq i \mid y_p = y_i\}$ denotes same-label positives of subject i within a minibatch, $\text{sim}(\cdot,\cdot)$ denotes cosine similarity, τ is the temperature, and β controls the reference strength. We set $w_{\text{ref}} = w_{\text{usl}} = 1.0$, $\tau = 0.1$, and $\beta = 0.65$ on the validation set.

2.5 Loss and Training

We first pretrain APP with $\mathcal{L}_{\text{APP}}$ to infer state-consistent phase boundaries via changepoint detection, and then optimize the remaining modules end-to-end with the following objective:

$$\mathcal{L} = \min \underbrace{\mathcal{L}_{\text{ce}} + \lambda_{\text{str}}\mathcal{L}_{\text{str}}}_{\substack{\text{Diagnosis-label}\\ \text{supervision}}} + \underbrace{\lambda_{\text{bin}}\mathcal{L}_{\text{bin}} + \lambda_{\text{ms}}\mathcal{L}_{\text{ms}} + \lambda_{\text{sp}}\mathcal{L}_{\text{sp}}}_{\text{Structure regularization}}, \tag{5}$$

where $\mathcal{L}_{\text{ce}}$ is the cross-entropy loss. We set $\lambda_{\text{str}} = 1.0$, $\lambda_{\text{bin}} = 1.0$, $\lambda_{\text{ms}} = 1.0$, and $\lambda_{\text{sp}} = 0.27$, determined via grid search on the validation set.

3 Experiments and Results

3.1 Dataset and Experimental Details

We evaluate BrainSTR on two rs-fMRI benchmarks for three diagnoses (**MDD**, **BD**, **ASD**). The private cohort constructed from Nanjing Medical University (NMU)[7] includes **246** controls, **151** MDD, and **126** BD subjects, preprocessed via **DPABI** [22]. From the public **ABIDE** dataset, we use only the **NYU** site (**74** ASD, **98** controls) to reduce multi-center variability. We used AAL-based pre-processed functional data for evaluation. Five-fold cross-validation is applied, with hyperparameters tuned on internal validation set.

[7] Approved by the Medical Science Ethics Committee of Nanjing Medical University (Ref. 2020-KY027-01).

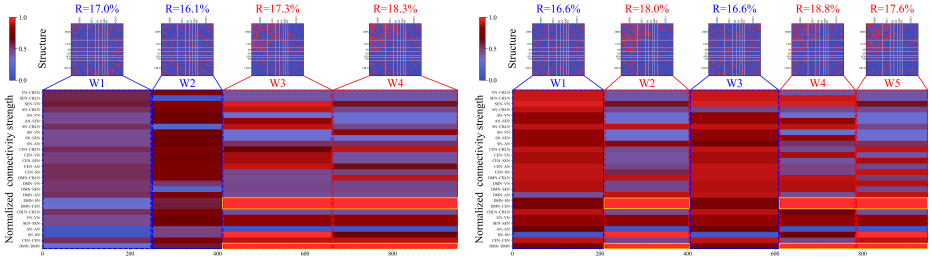
Table 1. Comparison with state-of-the-art methods on MDD, BD, and ASD diagnosis. The best results are in bold. $*$ indicates that BrainSTR significantly outperforms the compared method (paired t-test, $p<0.05$) across five folds.

Category	Method	HC vs. MDD		HC vs. BD		HC vs. ASD (NYU center)	
		ACC(%)	AUC(%)	ACC(%)	AUC(%)	ACC(%)	AUC(%)
Traditional	FC + SVM	$68.4 \pm 2.7^*$	$67.2 \pm 3.2^*$	$65.6 \pm 3.1^*$	$67.3 \pm 3.5^*$	$68.1 \pm 2.9^*$	$69.7 \pm 3.6^*$
	FC + RF	$64.7 \pm 2.9^*$	$61.2 \pm 3.4^*$	$65.2 \pm 3.1^*$	$63.6 \pm 4.0^*$	$66.2 \pm 2.8^*$	$65.1 \pm 3.7^*$
	SVM-RFE [3]	$70.3 \pm 3.8^*$	$70.1 \pm 4.0^*$	$66.9 \pm 3.7^*$	$67.6 \pm 3.9^*$	$65.2 \pm 4.1^*$	$65.9 \pm 4.3^*$
Static graph	GroupINN [23]	$67.1 \pm 3.9^*$	$65.7 \pm 4.2^*$	$68.3 \pm 3.8^*$	$65.2 \pm 4.0^*$	$63.3 \pm 4.3^*$	$64.3 \pm 4.4^*$
	BrainGNN [10]	$65.2 \pm 3.6^*$	$63.7 \pm 3.9^*$	$69.1 \pm 3.4^*$	$68.7 \pm 3.6^*$	$65.2 \pm 3.8^*$	$65.4 \pm 4.0^*$
	MVS-GCN [19]	$68.1 \pm 3.5^*$	$68.1 \pm 3.7^*$	$67.1 \pm 3.6^*$	$65.0 \pm 3.8^*$	69.3 ± 3.7	$68.3 \pm 3.9^*$
	ASD-DiagNet [1]	$68.2 \pm 3.4^*$	$68.0 \pm 3.6^*$	$72.1 \pm 3.2^*$	$72.1 \pm 3.4^*$	$67.4 \pm 3.5^*$	$67.5 \pm 3.7^*$
	FRL [8]	$68.0 \pm 3.3^*$	$68.9 \pm 3.5^*$	$71.1 \pm 3.2^*$	$70.7 \pm 4.1^*$	$67.1 \pm 3.5^*$	69.4 ± 3.7
	EAG-RS [6]	$66.5 \pm 3.6^*$	$70.9 \pm 3.8^*$	$73.5 \pm 3.1^*$	$72.1 \pm 4.3^*$	$66.8 \pm 3.7^*$	$68.1 \pm 3.9^*$
	D-CoRP [4]	$71.7 \pm 2.7^*$	$71.0 \pm 2.9^*$	$72.6 \pm 2.7^*$	$71.8 \pm 2.9^*$	$67.9 \pm 1.7^*$	$65.3 \pm 2.1^*$
Dynamic graph	MDGL [12]	$67.4 \pm 3.7^*$	$66.7 \pm 3.9^*$	$69.1 \pm 3.5^*$	$70.3 \pm 3.7^*$	$66.9 \pm 3.9^*$	$64.1 \pm 4.1^*$
	STA-GIN [9]	$73.1 \pm 3.2^*$	$66.0 \pm 3.5^*$	$72.4 \pm 3.1^*$	$68.5 \pm 3.3^*$	68.9 ± 3.4	$67.2 \pm 3.7^*$
	BrainDGT [14]	$74.3 \pm 2.6^*$	$71.3 \pm 2.8^*$	$73.7 \pm 3.4^*$	$74.7 \pm 2.7^*$	$69.1 \pm 3.1^*$	69.7 ± 4.2
	MCDGLN [18]	$72.6 \pm 2.5^*$	$71.1 \pm 3.7^*$	$71.6 \pm 3.2^*$	$72.1 \pm 2.6^*$	$67.2 \pm 2.7^*$	$69.3 \pm 2.9^*$
	ST-GCN [2]	$60.3 \pm 4.8^*$	$62.9 \pm 5.1^*$	$61.3 \pm 4.7^*$	$60.9 \pm 4.9^*$	$56.7 \pm 5.2^*$	$52.6 \pm 5.6^*$
Ours	w/o APP	74.3 ± 3.3	77.2 ± 3.5	75.9 ± 3.2	76.2 ± 3.4	70.1 ± 3.5	70.4 ± 3.7
	w/o ΔS	68.6 ± 4.4	64.3 ± 4.7	69.4 ± 4.3	66.8 ± 4.6	69.4 ± 4.6	66.8 ± 4.9
	w/o L_{str}	72.4 ± 3.8	77.3 ± 4.0	72.4 ± 3.7	74.5 ± 3.9	71.3 ± 4.0	72.7 ± 4.2
	w/o L_{ms}	74.3 ± 3.6	71.3 ± 3.8	73.7 ± 3.5	70.2 ± 3.7	71.7 ± 3.8	70.2 ± 4.0
	w/o L_{bin}	76.7 ± 3.1	75.2 ± 3.3	77.6 ± 3.0	78.2 ± 3.2	72.1 ± 3.3	71.8 ± 3.5
	w/o L_{sp}	76.2 ± 3.0	76.7 ± 3.2	77.8 ± 2.9	79.0 ± 3.1	72.1 ± 3.2	72.5 ± 3.4
Full Model	**BrainSTR**	**77.2±1.6**	**77.8±2.8**	**78.2±1.5**	**79.6±1.7**	**72.4±1.9**	**73.0±2.1**

3.2 Classification Results

Comparison with state-of-the-art. Table 1 compares BrainSTR with traditional classifiers built on handcrafted FC features, static-graph models on aggregated FC, and dynamic-graph models for dFC, where **BrainDGT** is the strongest dynamic baseline. On the NMU cohort, BrainSTR achieves $77.2 \pm 1.6/77.8 \pm 2.8$ (ACC/AUC, in %) for MDD and $78.2 \pm 1.5/79.6 \pm 1.7$ for BD, surpassing BrainDGT by $+2.9\%/+6.5\%$ and $+4.5\%/+4.9\%$, respectively. On ABIDE (NYU), BrainSTR remains the top-performing method. Overall, BrainSTR achieves the best performance in terms of both ACC and AUC across all three diseases, indicating robust cross-task superiority and generalization. We attribute the improvements to BrainSTR's explicit extraction of diagnosis-relevant spatio-temporal patterns: It learns stable brain-state trajectories via Adaptive Phase Partition (APP), incrementally estimates phase-wise structures to separate disease-related from irrelevant connectivity, and refines discriminative representations through spatio-temporal contrastive learning. Paired t-tests confirm that these gains are statistically significant ($p<0.05$) against the vast majority of baselines (Table 1).

Ablation analysis. We further verify the contribution of each key component in Table 1. Removing **APP** degrades performance consistently, indicating that adaptive boundary learning provides a better temporal partition for downstream FC construction and diagnosis. Discarding the transition cue ΔS causes



(a) Representative subject #1 (b) Representative subject #2

Fig. 2. Case-level temporal–topological interpretability of BrainSTR on representative subjects. Top: phase-wise structures, together with the overall retained-connectivity ratio R of each phase. Bottom: normalized connectivity strength of subnetworks.

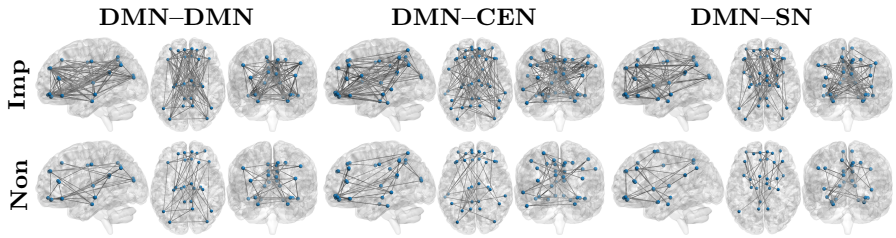


Fig. 3. Group-level DMN-centered structure retention. The top row (Imp) and bottom row (Non) visualize the group-aggregated retained connectivities in important and non-important phases, respectively, shown from three brain views.

the largest drop, highlighting that modeling incremental structure is crucial for capturing diagnosis-relevant dynamic connectivity. In addition, removing $\mathcal{L}_{\text{str}}$ or individual structure regularizers also leads to noticeable degradation, showing that these objectives are complementary in suppressing redundant connectivity patterns and promoting diagnosis-relevant spatio-temporal evidence.

3.3 Interpretability

Individual-level. On the MDD cohort, important phases (red) show higher R with denser DMN–DMN/DMN–CEN/DMN–SN retention than non-important phases (blue) (Fig. 2). The normalized connectivity strength highlights these DMN-related pairs in important phases (yellow). Temporally, the learned phase boundaries isolate state-consistent cognitive transitions, not transient noise; topologically, the retained triple-network (DMN, CEN, SN) interactions align with known disrupted dynamics in depression [13, 21, 16, 7].

Population-level. On the MDD cohort, the retained structures are denser in important phases, while non-important phases remain sparse and diffuse (Fig. 3). The BD cohort and ABIDE (NYU) show similar trends, supporting that BrainSTR captures diagnosis-relevant evidence both temporally and topologically.

4 Conclusion

In this work, we presented BrainSTR for interpretable dynamic brain-network diagnosis. BrainSTR combines Adaptive Phase Partition, critical-phase selection, structure-based topological filtering, and reference-adjusted contrastive supervision, shifting the focus from whole-sample aggregation to diagnosis-relevant spatio-temporal evidence. Experiments on ASD, BD, and MDD show consistent improvements over strong baselines. The identified critical phases and subnetworks are neurobiologically plausible, supporting the clinical interpretability of the proposed framework.

Acknowledgments. This research was supported by the National Natural Science Foundation of China (62076059), the Science and Technology Joint Project of Liaoning Province (2023JH2/101700367 and ZX20240193) and the Fundamental Research Funds for the Central Universities (N25BJD005).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Eslami, T., Mirjalili, V., Fong, A., Laird, A.R., Saeed, F.: Asd-diagnet: a hybrid learning approach for detection of autism spectrum disorder using fmri data. Frontiers in neuroinformatics **13**, 70 (2019)
2. Gadgil, S., Zhao, Q., Pfefferbaum, A., Sullivan, E.V., Adeli, E., Pohl, K.M.: Spatio-temporal graph convolution for resting-state fmri analysis. In: International conference on medical image computing and computer-assisted intervention. pp. 528–538. Springer (2020)
3. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine learning **46**(1), 389–422 (2002)
4. Hu, H., Zhang, H., Li, C.: D-corp: Differentiable connectivity refinement for functional brain networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 119–129. Springer (2024)
5. Jiang, N., Shi, Y.: A new adaptive sliding window method for fmri dynamic functional connectivity analysis. International Journal of Imaging Systems and Technology **35**(4), e70154 (2025)
6. Jung, W., Jeon, E., Kang, E., Suk, H.I.: Eag-rs: a novel explainability-guided roi-selection framework for asd diagnosis via inter-regional relation learning. IEEE Transactions on Medical Imaging **43**(4), 1400–1411 (2023)
7. Kaiser, R.H., Andrews-Hanna, J.R., Wager, T.D., Pizzagalli, D.A.: Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. JAMA psychiatry **72**(6), 603–611 (2015)
8. Kang, E., Heo, D.W., Lee, J., Suk, H.I.: A learnable counter-condition analysis framework for functional connectivity-based neurological disorder diagnosis. IEEE Transactions on Medical Imaging **43**(4), 1377–1387 (2023)
9. Kim, B.H., Ye, J.C., Kim, J.J.: Learning dynamic graph representation of brain connectome with spatio-temporal attention. Advances in Neural Information Processing Systems **34**, 4314–4327 (2021)

10. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L.H., Ventola, P., Duncan, J.S.: Braingnn: Interpretable brain graph neural network for fmri analysis. Medical Image Analysis **74**, 102233 (2021)
11. Liu, L., Wen, G., Cao, P., Hong, T., Yang, J., Zhang, X., Zaiane, O.R.: Braintgl: A dynamic graph representation learning model for brain network analysis. Computers in Biology and Medicine **153**, 106521 (2023)
12. Ma, Y., Wang, Q., Cao, L., Li, L., Zhang, C., Qiao, L., Liu, M.: Multi-scale dynamic graph learning for brain disorder detection with functional mri. IEEE Transactions on Neural Systems and Rehabilitation Engineering **31**, 3501–3512 (2023)
13. Manoliu, A., Meng, C., Brandl, F., Doll, A., Tahmasian, M., Scherr, M., Schwerthöffer, D., Zimmer, C., Förstl, H., Bäuml, J., et al.: Insular dysfunction within the salience network is associated with severity of symptoms and aberrant inter-network connectivity in major depressive disorder. Frontiers in human neuroscience **7**, 930 (2014)
14. Shehzad, A., Zhang, D., Yu, S., Abid, S., Xia, F.: Dynamic graph transformer for brain disorder diagnosis. IEEE Journal of Biomedical and Health Informatics (2025)
15. Šverko, Z., Vrankic, M., Vlahinić, S., Rogelj, P.: Dynamic connectivity analysis using adaptive window size. Sensors **22**(14), 5162 (2022)
16. Tozzi, L., Zhang, X., Chesnut, M., Holt-Gosselin, B., Ramirez, C.A., Williams, L.M.: Reduced functional connectivity of default mode network subsystems in depression: Meta-analytic evidence and relationship with trait rumination. NeuroImage: Clinical **30**, 102570 (2021)
17. Wang, N., Yao, D., Ma, L., Liu, M.: Multi-site clustering and nested feature extraction for identifying autism spectrum disorder with resting-state fmri. Medical image analysis **75**, 102279 (2022)
18. Wang, P., Wen, X., Lei, Y., Guo, Y., Li, J., Hao, Y., Cao, R., Gao, C., Cao, R.: Mcdgln: Masked connection-based dynamic graph learning network for autism spectrum disorder. Brain Research Bulletin **224**, 111290 (2025)
19. Wen, G., Cao, P., Bao, H., Yang, W., Zheng, T., Zaiane, O.: Mvs-gcn: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. Computers in biology and medicine **142**, 105239 (2022)
20. Wen, G., Cao, P., Liu, L., Hao, M., Liu, S., Zheng, J., Yang, J., Zaiane, O.R., Wang, F.: Heterogeneous graph representation learning framework for resting-state functional connectivity analysis. IEEE Transactions on Medical Imaging (2024)
21. Yan, C.G., Chen, X., Li, L., Castellanos, F.X., Bai, T.J., Bo, Q.J., Cao, J., Chen, G.M., Chen, N.X., Chen, W., et al.: Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. Proceedings of the National Academy of Sciences **116**(18), 9078–9083 (2019)
22. Yan, C.G., Wang, X.D., Zuo, X.N., Zang, Y.F.: Dpabi: data processing & analysis for (resting-state) brain imaging. Neuroinformatics **14**, 339–351 (2016)
23. Yan, Y., Zhu, J., Duda, M., Solarz, E., Sripada, C., Koutra, D.: Groupinn: Grouping-based interpretable neural network for classification of limited, noisy brain data. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 772–782 (2019)
24. Zhang, P., Wen, G., Cao, P., Yang, J., Zhang, J., Zhang, X., Zhu, X., Zaiane, O.R., Wang, F.: Brainusl: Unsupervised graph structure learning for functional brain network analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 205–214. Springer (2023)