Spectral-Temporal State Space Modeling on Functional Brain Networks

Jaeyoon Sim[1*], Hoseok Lee[1*], Jihwan Park[1], Seunghun Baek[1], Yu Zhang[2], and Won Hwa Kim[1]

[1] Pohang University of Science and Technology, Pohang, South Korea
{simjy98, hslee0608, pjh58110, habaek4, wonhwa}@postech.ac.kr
[2] Stanford University, Palo Alto CA, USA

Abstract. Resting-state fMRI (rs-fMRI) offers a powerful tool for analyzing functional organization in brain for neurodegenerative and neurodevelopmental disorders. Although graph-based and spatio-temporal models have shown promise, most existing approaches decouple spatial structure from temporal dynamics or rely on predefined temporal windows, limiting the capacity of spatial information to directly modulate temporal processing. To address these limitations, we propose a spectral–temporal state space framework that integrates graph spectral representations with state-space modeling through a spectral-temporal kernel, enabling window-free and frequency-based temporal modeling. Experiments on three rs-fMRI cohorts with 1,926 subjects demonstrate that our method consistently outperforms competitive baselines across diverse conditions, including Parkinson's disease, autism spectrum disorder, and attention deficit hyperactivity disorder. In addition, we provide interpretable insights into spectral–temporal patterns of brain dysfunction, advancing the characterization and classification of neurodegenerative and neurodevelopmental disorders.

Keywords: Graph Spectral Analysis · State-Space Modeling · Brain Network Analysis · rs-fMRI Disease Classification.

1 Introduction

Functional magnetic resonance imaging (fMRI) is a *non-invasive* neuroimaging modality that measures blood-oxygen-level-dependent (BOLD) signals to characterize spatio-temporal brain activity and functional connectivity. These capabilities have led to the widespread adoption of fMRI in both clinical and research settings for studying neurodevelopmental and neurodegenerative disorders, including Parkinson's disease, autism spectrum disorder, and attention deficit hyperactivity disorder [22,25,34]. In practice, correlations among BOLD time series extracted in predefined regions of interest (ROIs) construct functional brain networks, which allow the investigation of inter-regional coupling patterns [7,9]. Building on this representation, deep learning approaches have increasingly

[*] J. Sim and H. Lee contributed equally to this paper.

been explored to model fMRI-derived brain networks for data-driven prediction and characterization of neurological and psychiatric conditions.

Existing methods for brain network analysis aim to identify neurobiological markers by decoding the complex topological structures of brain function [27,28]. For example, BrainGNN [17] introduces pooling operators to identify critical nodes, BrainNetTF [11] incorporates orthogonal clustering into the graph transformer to obtain discriminative node-level embeddings, ALTER [38] captures long-range dependencies through biased random walks, BioBGT [24] proposes entanglement-based node importance encoding, and BQN [37] adaptively learns brain functional networks using a quadratic network. However, these static approaches mainly rely on time-averaged connectivity and therefore struggle to capture the transient and non-stationary properties of neural dynamics.

To address this limitation, spatio-temporal approaches incorporate spatial topology with temporal fluctuations by modeling the brain as a dynamic system. For instance, STAGIN [12] extracts structural features with temporal dynamics by attention mechanisms, BolT [1] captures spatio-temporal information using fused window attention, BrainWaveNet [10] models temporal dynamics via continuous wavelet transformation, STG-Mamba [15] integrates graph information with state-space temporal modeling, and FST-Mamba [32] hierarchically captures spatio-temporal information. Despite their effectiveness, most recent approaches still process topology and dynamics in a decoupled manner, fusing them only as a post-processing step, which restricts topology-driven temporal modeling. Moreover, the reliance on sliding windows often renders performance sensitive to window choices.

In this work, we propose **Spec**tral-**T**emporal **Mamba** (i.e., SpecT-Mamba) that directly couples graph spectral structure with state-space temporal modeling via a learnable spectral–temporal kernel. This kernel adaptively modulates signal amplitudes to prioritize informative frequencies while simultaneously adjusting memory scales to control the temporal decay rate, allowing for the capture of both transient and sustained neural activities. By integrating these topology-guided dynamics rather than relying on post-hoc fusion, SpecT-Mamba captures multi-scale temporal patterns in fMRI signals without relying on temporal partitioning such as sliding windows. As a result, we reformulate the proposed spectral-temporal filtering as a state-space model, achieving strong downstream performance while providing insights into spectral-based temporal patterns.

The key contributions of our work are **1)** introducing a novel spectral–temporal state space framework built upon a learnable spectral–temporal kernel that jointly modulates temporal amplitudes and memory supports of signals, **2)** enabling window-free spatio-temporal modeling without sensitivity to heuristic hyperparameter choices, and **3)** demonstrating superior performance on rs-fMRI benchmarks while offering insight into temporal dynamics. Experiments on publicly available rs-fMRI benchmarks demonstrate that the proposed framework achieves strong performance in classifying and interpreting brain diseases, highlighting its potential to support early diagnosis in clinical settings.

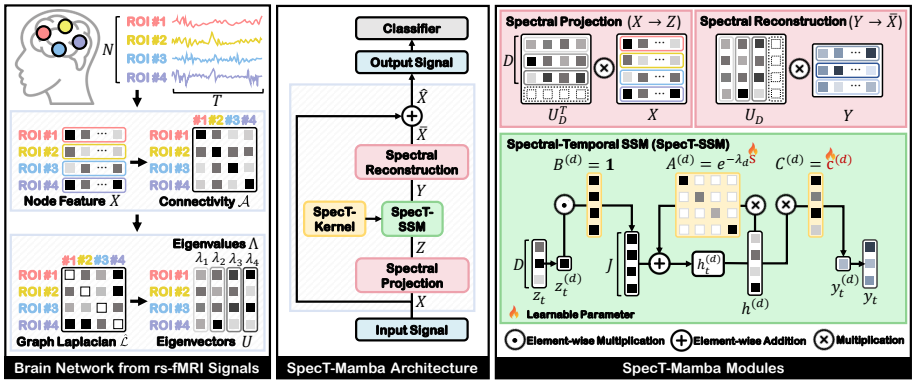

Fig. 1: Illustration of SpecT-Mamba. Given an input signal X, the X is projected to the graph spectral subspace as Z and processed by the spectral-temporal SSM (SpecT-SSM), where the spectral output Y are obtained through a spectral-temporal kernel (SpecT-Kernel) by optimizing parameters c and s. Then, the Y is reconstructed to the graph space as $\bar{X}$ and fed into the downstream classifier.

2 Preliminary: State Space Model

State Space Models (SSM) [6] are sequence models that represent temporal processes using state equations describing the evolution of latent states. For an input $x(t) \in \mathbb{R}$, a hidden state $h(t) \in \mathbb{R}^H$ with state dimension H, and an output $y(t) \in \mathbb{R}$, the continuous-time state equation is expressed as

$$h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t), \quad y(t) = \mathbf{C}h(t). \tag{1}$$

Here, $\mathbf{A} \in \mathbb{R}^{H\times H}$ is the state matrix, $\mathbf{B} \in \mathbb{R}^H$ is the input projection matrix, and $\mathbf{C} \in \mathbb{R}^{1\times H}$ is the output projection matrix. While these state equations are typically defined in continuous time to capture real-world dynamics, Mamba [5] enables end-to-end learning by discretizing Eq. (1) via Zero-Order Hold (ZOH) [23]. With the time scaling parameter $\Delta \in \mathbb{R}$ from ZOH discretization, the continuous-time parameters are mapped into discrete matrices $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ as

$$\bar{\mathbf{A}} = \exp(\Delta\mathbf{A}), \quad \bar{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I}) \cdot \Delta\mathbf{B} \approx \Delta\mathbf{B}, \tag{2}$$

where $\mathbf{A}$ is restricted to be diagonal for stable training under ZOH discretization. When A is diagonal and Δ is small, the ZOH input term is commonly approximated as $\bar{B} \approx \Delta B$. Based on Eq. (2), Eq. (1) is converted as

$$h_t = \bar{\mathbf{A}}h_{t-1} + \bar{\mathbf{B}}x_t, \quad y_t = \mathbf{C}h_t. \tag{3}$$

3 Spectral-Temporal Mamba (SpecT-Mamba)

Brain Network from rs-fMRI Signals. A brain network is represented as an undirected graph $\mathcal{G}=\{\mathcal{V}, \mathcal{E}\}$ with a node set $\mathcal{V}$ and an edge set $\mathcal{E}$, where each

node corresponds to a ROI. As in Fig. 1, the rs-fMRI signal is defined as a node feature matrix $X \in \mathbb{R}^{N\times T}$ with N ROIs, where each ROI is associated with a BOLD time series of length T. To construct a functional connectivity matrix $\mathcal{A}$, Pearson correlation coefficients [12] are computed between ROI time series, and the resulting correlation matrix is then thresholded at 0.5 to obtain a binary adjacency matrix $\hat{\mathcal{A}} \in \mathbb{R}^{N\times N}$. Based on $\hat{\mathcal{A}}$, the graph Laplacian is defined as $\mathcal{L}{=}\mathcal{D}{-}\hat{\mathcal{A}}$, where $\mathcal{D} \in \mathbb{R}^{N\times N}$ denotes the diagonal degree matrix. It has a complete set of orthonormal eigenvectors $U{=}[u_1|u_2|\ldots|u_N]$ with corresponding real and non-negative eigenvalues $0{=}\lambda_1{\leq}\lambda_2{\leq}\ldots{\leq}\lambda_N$, so does the normalized Laplacian $\hat{\mathcal{L}}{=}\mathcal{D}^{-1/2}\mathcal{L}\mathcal{D}^{-1/2}$.

Spectral Projection. Using $\hat{\mathcal{L}}$, X is projected onto U to obtain a graph Fourier representation. This transformation decomposes node signals into frequency components, which enables explicit modeling of dynamics across different spatial frequencies. By selecting the D lowest eigenvalues $\{\lambda_d\}_{d=1}^{D}$, the representation emphasizes smooth spectral components while suppressing high-frequency noise from raw signals. Formally, this transformation is given by $Z{=}U_D^\top X$, where $Z \in \mathbb{R}^{D\times T}$ denotes the spectral coefficients of the rs-fMRI and $U_D \in \mathbb{R}^{N\times D}$ is the truncated eigenbasis corresponding to the lowest D eigenvalues. For each spectral channel d, the spectral coefficient at time t is defined as $z_t^{(d)}{=}u_d^\top x_t$, where $u_d \in \mathbb{R}^N$ is the d-th eigenvector and $x_t \in \mathbb{R}^N$ is the t-th node signal.

Spectral–Temporal Kernel. After projection, the spectral output $y_t \in \mathbb{R}^D$ at time t is obtained by aggregating past coefficients as

$$y_t = \sum_{\tau\leq t} g_{t,\tau}(\Lambda_D) z_\tau = \sum_{\tau\leq t} g_{t,\tau}(\Lambda_D)\, U_D^\top x_\tau, \tag{4}$$

where $\Lambda_D = \mathrm{diag}[\lambda_1,\ldots,\lambda_D]$ and $g_{t,\tau}(\cdot)$ is a spectral-temporal kernel that maps the spectral coefficient at time τ to the output at time t. To model multi-scale temporal dependencies in a topology-aware manner, we design this kernel that jointly modulates (i) the amplitude of spectral components and (ii) the frequency-conditioned temporal decay rate. For each d, we parameterize $g_{t,\tau}^{(d)}$ as a mixture of exponential kernels over J different temporal scales, given by

$$g_{t,\tau}^{(d)} = \sum_{j=1}^{J} c_j^{(d)} \cdot e^{a_j^{(d)}(t-\tau)} = \sum_{j=1}^{J} c_j^{(d)} \cdot e^{-s_j\lambda_d(t-\tau)}. \tag{5}$$

Here, the learnable coefficient $c_j^{(d)}$ modulates the amplitude contribution of the j-th scale in the d-th spectral component, while the decay rate $a_j^{(d)}{=}{-}s_j\lambda_d$ couples a trainable and non-negative memory scale s_j with λ_d to control the decay speed. This parameterization allows SpecT-Mamba to adaptively emphasize or suppress spectral strength and effective temporal support in a frequency-dependent manner. Therefore, by jointly optimizing $c_j^{(d)}$ and s_j, SpecT-Mamba adaptively reweights temporal influences and adjusts their memory spans to extract the most discriminative spectral–temporal patterns for disease classification.

Spectral-Temporal SSM. Building on the spectral-temporal kernel in Eq. (5), we express the induced causal filtering in a state-space form. For spectral channel

d, filtering the spectral coefficients $z_t^{(d)}$ at time t produces $y_t^{(d)}=\sum_{\tau\le t} g_{t,\tau}^{(d)} \cdot z_\tau^{(d)}$. Since $g_{t,\tau}^{(d)}$ is parameterized as a sum of exponentials, a scale-wise hidden representation $h_{t,s_j}^{(d)} \in \mathbb{R}$ of the d-th spectral component at time t is defined as

$$\begin{aligned} h_{t,s_j}^{(d)} &= \sum_{\tau\le t} e^{-s_j\lambda_d(t-\tau)} \cdot z_\tau^{(d)} = \sum_{\tau\le t-1} e^{-s_j\lambda_d(t-\tau)} \cdot z_\tau^{(d)} + z_t^{(d)} \\ &= e^{-s_j\lambda_d} \sum_{\tau\le t-1} e^{-s_j\lambda_d(t-1-\tau)} \cdot z_\tau^{(d)} + z_t^{(d)} = e^{-s_j\lambda_d} \cdot h_{t-1,s_j}^{(d)} + z_t^{(d)}, \end{aligned} \tag{6}$$

which yields the spectral output $y_t^{(d)} \in \mathbb{R}$ over J memory scales as

$$y_t^{(d)} = \sum_{\tau\le t} g_{t,\tau}^{(d)} \cdot z_\tau^{(d)} = \sum_{j=1}^{J} c_j^{(d)} \cdot h_{t,s_j}^{(d)}. \tag{7}$$

By defining $\mathbf{A}^{(d)} = \mathrm{diag}[e^{-s_1\lambda_d}, \ldots, e^{-s_J\lambda_d}] \in \mathbb{R}^{J\times J}$, $\mathbf{B}^{(d)} = \mathbf{1}_J \in \mathbb{R}^J$, and $\mathbf{C}^{(d)} = [c_1^{(d)}, \ldots, c_J^{(d)}] \in \mathbb{R}^{1\times J}$, Eq. (6) and (7) align with the canonical discrete-time SSM as Eq. (3). Consequently, the proposed kernel is implemented as a Mamba block, achieving efficient sequential modeling while preserving the intrinsic spectral–temporal structures of brain networks.

Reconstruction. To recover node-level representations in the graph space, the spectral output $y_t \in \mathbb{R}^D$ is transformed back to the graph space using the U_D, yielding the reconstructed node signal $\bar{x}_t \in \mathbb{R}^N$ at time t as

$$\bar{x}_t = U_D y_t = U_D \sum_{\tau\le t} g_{t,\tau}(\Lambda_D)\, U_D^\top x_\tau. \tag{8}$$

Finally, the reconstructed node feature $\bar{X} \in \mathbb{R}^{N\times T}$ is combined with the original input X through a weighted skip connection controlled by α, i.e., $\hat{X} = \alpha\bar{X} + (1-\alpha)X$, which is fed into the downstream classifier for disease prediction.

4 Experiments

Datasets. To evaluate the performance of SpecT-Mamba, we utilized public rs-fMRI datasets; *Parkinson's Progression Markers Initiative (PPMI)* [20] as a neurodegenerative benchmark, alongside *Autism Brain Imaging Data Exchange (ABIDE)* [8] and *Attention Deficit Hyperactivity Disorder (ADHD-200)* [2] as neurodevelopmental benchmarks. Following [36] and [2], fMRI scans from N=181 (68 Non-PD and 113 PD) subjects in PPMI, N=989 (534 Control and 455 Autism) in ABIDE, and N=776 (488 Control and 268 ADHD) in ADHD-200 are transformed into brain networks. Each brain is parcellated into 116 regions using the AAL atlas [29]. To mitigate site-driven differences in sequence length and maintain temporal consistency across subjects, we truncated BOLD signals to the shortest sequence within each dataset. Thus, the temporal dimensions were set to 112, 120, and 72 for PPMI, ABIDE, and ADHD-200, respectively.

Table 1: Comparison of classification performance between static (top) and spatio-temporal (bottom) methods on the three rs-fMRI brain network benchmarks. The best , second-best , and third-best results are highlighted.

Method	PPMI			ABIDE			ADHD-200		
	Acc ↑	Pre ↑	Rec ↑	Acc ↑	Pre ↑	Rec ↑	Acc ↑	Pre ↑	Rec ↑
GCN [14]	$55.2_{\pm 7.0}$	$52.7_{\pm 6.6}$	$52.5_{\pm 6.5}$	$50.5_{\pm 3.2}$	$50.3_{\pm 3.4}$	$50.3_{\pm 3.4}$	$54.8_{\pm 4.9}$	$52.0_{\pm 4.4}$	$52.0_{\pm 4.5}$
GAT [30]	$56.9_{\pm 8.8}$	$55.3_{\pm 7.9}$	$54.4_{\pm 6.9}$	$53.2_{\pm 2.2}$	$53.0_{\pm 2.5}$	$53.0_{\pm 2.4}$	$55.2_{\pm 3.5}$	$53.0_{\pm 1.9}$	$53.0_{\pm 1.8}$
BrainGNN [17]	$59.1_{\pm 5.4}$	$57.4_{\pm 5.7}$	$57.8_{\pm 6.2}$	$55.7_{\pm 1.5}$	$58.9_{\pm 6.2}$	$52.9_{\pm 2.1}$	$58.2_{\pm 3.4}$	$57.1_{\pm 3.2}$	$57.2_{\pm 3.4}$
BrainNetTF [11]	$60.2_{\pm 5.2}$	$62.2_{\pm 7.4}$	$60.2_{\pm 4.9}$	$53.0_{\pm 2.9}$	$55.7_{\pm 2.5}$	$53.8_{\pm 1.9}$	$55.2_{\pm 4.4}$	$50.9_{\pm 4.2}$	$50.7_{\pm 3.9}$
ALTER [38]	$62.9_{\pm 8.2}$	$60.5_{\pm 8.1}$	$60.7_{\pm 8.0}$	$61.9_{\pm 2.6}$	$61.8_{\pm 2.6}$	$61.7_{\pm 2.4}$	$57.4_{\pm 2.9}$	$54.1_{\pm 2.7}$	$53.9_{\pm 2.1}$
Graph-Mamba [31]	$61.9_{\pm 6.0}$	$61.6_{\pm 6.3}$	$62.2_{\pm 6.4}$	$56.0_{\pm 2.4}$	$49.3_{\pm 9.6}$	$53.8_{\pm 1.9}$	$51.3_{\pm 3.7}$	$54.7_{\pm 3.5}$	$54.7_{\pm 3.5}$
BioBGT [24]	$61.3_{\pm 5.1}$	$60.8_{\pm 4.9}$	$59.0_{\pm 5.1}$	$56.3_{\pm 1.4}$	$55.9_{\pm 1.5}$	$55.0_{\pm 1.5}$	$60.2_{\pm 3.2}$	$58.3_{\pm 6.7}$	$57.9_{\pm 7.1}$
BQN [37]	$64.6_{\pm 3.8}$	$62.9_{\pm 7.0}$	$59.6_{\pm 7.6}$	$54.5_{\pm 2.8}$	$55.7_{\pm 3.5}$	$54.9_{\pm 3.0}$	$61.5_{\pm 4.6}$	$59.2_{\pm 3.4}$	$56.1_{\pm 3.8}$
STAGIN-SERO [12]	$63.0_{\pm 6.8}$	$61.0_{\pm 8.3}$	$59.9_{\pm 6.7}$	$56.1_{\pm 2.5}$	$55.3_{\pm 2.9}$	$54.4_{\pm 2.9}$	$62.7_{\pm 5.0}$	$58.6_{\pm 7.1}$	$54.6_{\pm 3.4}$
STAGIN-GARO [12]	$67.4_{\pm 5.5}$	$66.3_{\pm 5.2}$	$66.0_{\pm 4.3}$	$57.7_{\pm 2.6}$	$57.3_{\pm 2.9}$	$56.7_{\pm 3.5}$	$62.0_{\pm 2.5}$	$53.3_{\pm 6.6}$	$53.2_{\pm 4.2}$
BolT [1]	$64.6_{\pm 5.8}$	$66.3_{\pm 6.9}$	$64.6_{\pm 6.3}$	$62.1_{\pm 3.1}$	$62.6_{\pm 3.2}$	$61.5_{\pm 3.1}$	$59.2_{\pm 4.9}$	$57.5_{\pm 6.3}$	$55.6_{\pm 5.3}$
Mamba [5]	$66.9_{\pm 1.5}$	$65.2_{\pm 2.7}$	$64.7_{\pm 2.1}$	$60.9_{\pm 3.1}$	$60.6_{\pm 3.4}$	$60.5_{\pm 3.3}$	$59.8_{\pm 1.7}$	$56.5_{\pm 2.3}$	$56.5_{\pm 2.3}$
BrainWaveNet [10]	$66.8_{\pm 3.5}$	$62.5_{\pm 7.5}$	$61.2_{\pm 7.2}$	$62.2_{\pm 3.8}$	$62.1_{\pm 4.4}$	$60.4_{\pm 5.3}$	$63.4_{\pm 3.8}$	$53.2_{\pm 6.2}$	$45.8_{\pm 6.5}$
STG-Mamba [15]	$67.3_{\pm 7.0}$	$65.3_{\pm 7.7}$	$62.5_{\pm 7.1}$	$62.1_{\pm 1.6}$	$62.8_{\pm 1.5}$	$62.7_{\pm 1.2}$	$62.6_{\pm 3.0}$	$58.4_{\pm 3.4}$	$57.5_{\pm 3.0}$
FST-Mamba [32]	$67.9_{\pm 8.9}$	$65.2_{\pm 9.4}$	$65.0_{\pm 9.9}$	$60.9_{\pm 3.0}$	$62.0_{\pm 4.0}$	$58.8_{\pm 1.2}$	$61.0_{\pm 4.0}$	$57.7_{\pm 4.2}$	$57.4_{\pm 3.7}$
SpecT-Mamba (Ours)	$71.3_{\pm 2.7}$	$69.3_{\pm 3.9}$	$69.5_{\pm 4.2}$	$64.1_{\pm 1.6}$	$63.8_{\pm 1.4}$	$62.5_{\pm 2.3}$	$66.6_{\pm 3.0}$	$62.6_{\pm 3.7}$	$61.3_{\pm 3.8}$

Baselines. We categorized the baselines into two groups; 1) *Static* methods such as GCN [14], GAT [30], BrainGNN [17], BrainGB [3], BrainNetTF [11], ALTER [38], Graph-Mamba [31], BioBGT [24], and BQN [37] and 2) *Spatio-Temporal* methods such as STAGIN [12] with two modules (i.e., SERO and GARO), BolT [1], Mamba [5], BrainWaveNet [10], STG-Mamba [15], and FST-Mamba [32].

Experimental Setup. For unbiased results, 5-fold cross validation was performed, and accuracy (Acc), precision (Pre), and recall (Rec) in their mean are reported. All baselines were trained using the Focal loss [19] and the Adam optimizer [13]. To ensure fair comparisons, each baseline was tuned to achieve best feasible outcomes. For SpecT-Mamba, we used a dropout rate of 0.5, a learning rate of 5e-3, and a weight decay of 5e-4. Also, we set the number of spectral channels D=64, with the scales J set to 8/16/4 for PPMI/ABIDE/ADHD-200.

Classification Result. The performance comparisons between SpecT-Mamba and 15 baselines including static and spatio-temporal methods across three benchmarks are reported in Table 1. While static baselines show reasonable performance, their results remain noticeably lower than those of spatio-temporal methods including SpecT-Mamba. Specifically, SpecT-Mamba brings improvements with average gain of 3.4%p in accuracy, 2.0%p in precision, and 3.5%p in recall over the second-best method for the PPMI. Notably, the superior performance of SpecT-Mamba stems from its effectiveness in adaptively modulating multi-scale temporal memories based on the graph spectral structure.

5 Interpretation of trained SpecT-Mamba

Analysis on the Interpretable Results. To investigate the neurobiological relevance of the patterns learned by SpecT-Mamba, we utilize Grad-CAM [26] to

Table 2: The Grad-CAM [26] results describing the top-10 ROIs with the highest activation (Act.) to classify disease on the PPMI (Left) and ADHD-200 (Right) datasets. The indices align with the index values in the AAL116 atlas [29], and (L) and (R) denote the left and right hemispheres, respectively.

PPMI					
Idx	**ROI**	**Act.**	**Idx**	**ROI**	**Act.**
75	(L) Pallidum	1.000	25	(L) Frontal.Mid.Orb	0.613
69	(L) Paracentral.Lob	0.824	83	(L) Templ.Pole.Sup	0.595
81	(L) Temporal.Sup	0.795	100	(R) Cerebellum.6	0.568
107	(L) Cerebellum.10	0.678	26	(R) Frontal.Mid.Orb	0.564
28	(R) Rectus	0.622	19	(L) Supp.Motor	0.523

ADHD-200					
Idx	**ROI**	**Act.**	**Idx**	**ROI**	**Act.**
25	(L) Frontal.Mid.Orb	1.000	60	(R) Parietal.Sup	0.892
94	(R) Cerebellum.Crus2	0.980	18	(R) Rolandic.Oper	0.889
78	(R) Thalamus	0.955	21	(L) Olfactory	0.879
20	(R) Supp.Motor	0.934	32	(R) Cingulum.Ant	0.855
59	(L) Parietal.Sup	0.921	8	(R) Frontal.Mid	0.842

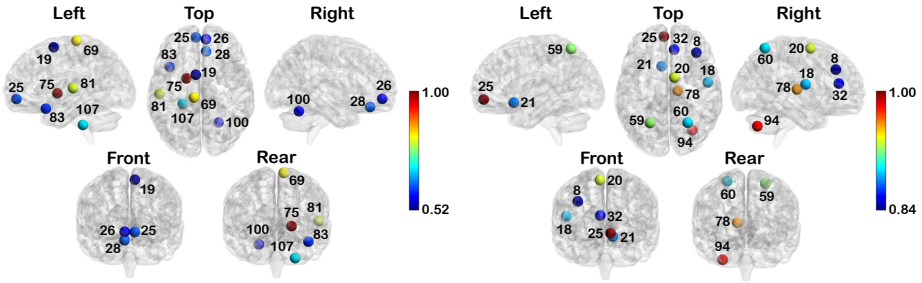

Fig. 2: Top-10 ROIs with the highest activation for classifying disease on the PPMI (Left) and ADHD-200 (Right). Node color indicates the activation.

identify ROIs that contribute most significantly to disease classification. Table 2 presents the top 10 nodes with the highest gradient activations for the *PPMI* and *ADHD-200* datasets, while their spatial distributions are visualized in Fig. 2.

For *PPMI*, the *left pallidum* shows the highest activation, with the *superior temporal gyrus* and *left cerebellum* also highlighted. These regions belong to the motor and default mode network, which are known to be disrupted in PD due to impaired motor execution and sensorimotor integration [16,33,35]. For *ADHD-200*, the *orbital part of the middle frontal gyrus* is the highest ranked ROI, and it is typically associated with the default mode network. In addition, regions within the salience network (e.g., *anterior cingulate cortex*) and frontoparietal control network (e.g., *cerebellar crus II*) are also highlighted, reflecting deficits in cognitive control and response inhibition in ADHD [4,18,21]. These results indicate that SpecT-Mamba captures both region-specific importance and structure-aware temporal dynamics through spectral–temporal modeling, supporting its relevance to neurological disorders.

Spectral-Temporal Dynamics via Kernels. To examine the temporal influence across spectral channels, Fig. 3 shows the kernel weights $g_{T,\tau}$ at the final step T over past times $\tau \leq T$ for each dataset. Here, the spectral channels are equally divided into low-, middle-, and high-frequency bands based on their eigenvalues. Across all cases, low-frequency groups assign larger weights to earlier time steps, indicating stronger long-range dependency, whereas high-frequency groups decay faster and emphasize recent inputs. This behavior reflects the distinct roles of

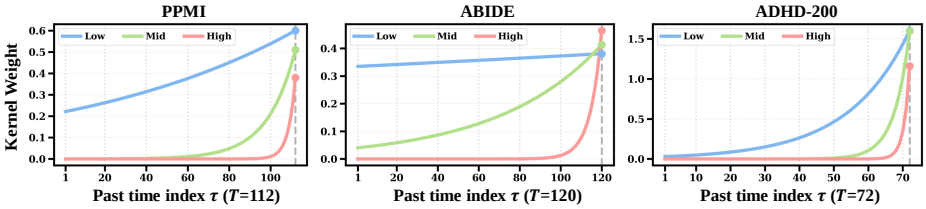

Fig. 3: Visualization of the learned spectral-temporal kernel weights $g_{T,\tau}$ on the three benchmarks. We fix the current time to the last step (i.e., $t{=}T$) and plot the kernel weight as a function of the past index τ, where spectral indices d are grouped into low-, middle-, and high-frequency bands.

Table 3: Ablation study on spectral-temporal coupling parameters of SpecT-Mamba.

Method	Memory Scale s	Amplitude Modulation c	PPMI			ADHD-200		
			Acc ↑	Pre ↑	Rec ↑	Acc ↑	Pre ↑	Rec ↑
SpecT-Mamba	Fixed	Fixed	$65.5_{\pm 3.8}$	$63.9_{\pm 4.8}$	$63.9_{\pm 5.1}$	$62.8_{\pm 2.1}$	$58.1_{\pm 2.5}$	$57.1_{\pm 2.4}$
	Fixed	Learnable	$67.9_{\pm 3.4}$	$67.1_{\pm 5.1}$	$65.7_{\pm 4.4}$	$63.9_{\pm 1.4}$	$59.6_{\pm 1.6}$	$58.9_{\pm 1.9}$
	Learnable	Fixed	$70.1_{\pm 5.7}$	$68.1_{\pm 5.7}$	$68.3_{\pm 6.3}$	$64.9_{\pm 2.4}$	$61.1_{\pm 2.1}$	$60.7_{\pm 2.3}$
	Learnable	Learnable	$\mathbf{71.3}_{\pm \mathbf{2.7}}$	$\mathbf{69.3}_{\pm \mathbf{3.9}}$	$\mathbf{69.5}_{\pm \mathbf{4.2}}$	$\mathbf{66.6}_{\pm \mathbf{3.0}}$	$\mathbf{62.6}_{\pm \mathbf{3.7}}$	$\mathbf{61.3}_{\pm \mathbf{3.8}}$

spectral components, where low-frequencies capture spatially smooth patterns that evolve slowly and thus benefit from long-range temporal integration, while high-frequencies encode localized and noise-sensitive variations for which short-range temporal modeling is appropriate to preserve transient dynamics. Overall, SpecT-Mamba adaptively allocates temporal influence according to spectral frequency, enabling differentiated temporal behaviors across spectral bands.

Effect of Spectral–Temporal Coupling Parameters. To investigate the influence of spectral–temporal coupling in SpecT-Mamba, we performed an ablation study on the memory scale s and the amplitude modulation coefficient c, as summarized in Table 3. When both s and c were fixed, performance showed the largest degradation, and learning either s or c led to a noticeable improvement. Performance was further boosted up to 71.3% on PPMI and 66.6% on ADHD-200 when both parameters were optimized jointly, demonstrating that capturing the spectral–temporal relationship adaptively is highly effective. This indicates that the performance gain mainly stems from jointly learning the spectral and temporal interactions, rather than specific hyperparameter settings.

6 Conclusion

In this work, we proposed SpecT-Mamba, a spectral–temporal framework that conditions temporal evolution on graph spectral structure. By modeling a multi-scale state-space kernel to modulate signal amplitude and temporal decay, our model captures both short- and long-range dependencies. Thus, SpecT-Mamba learns topology-aware temporal representations in which spectral channels encode dynamics relevant to neurological and neuropsychiatric disorders. Experi-

ments on rs-fMRI datasets show that SpecT-Mamba outperforms baselines while providing insights into spectral–temporal patterns of brain dysfunction.

Acknowledgments. This research was supported by RS-2026-25494850 (60%), RS-2025-02216257 (20%), GTL25071-100 (10%), and RS-2019-II1091906 (AI Graduate Program at POSTECH, 10%).

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

References

1. Bedel, H.A., Sivgin, I., Dalmaz, O., et al.: Bolt: Fused window transformers for fmri time series analysis. Medical Image Analysis (2023)
2. Brown, M.R., Sidhu, G.S., Greiner, R., et al.: Adhd-200 global competition: diagnosing adhd using personal characteristic data can outperform resting state fmri measurements. Frontiers in Systems Neuroscience (2012)
3. Cui, H., Dai, W., Zhu, Y., et al.: Braingb: a benchmark for brain network analysis with graph neural networks. Transactions on Medical Imaging (2022)
4. Depue, B.E., Burgess, G.C., Bidwell, L.C., et al.: Behavioral performance predicts grey matter reductions in the right inferior frontal gyrus in young adults with combined type adhd. Psychiatry Research: Neuroimaging (2010)
5. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. In: First conference on Language Modeling (2024)
6. Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. In: International Conference on Learning Representations (2022)
7. Han, S., Sun, Z., Zhao, K., et al.: Early prediction of dementia using fmri data with a graph convolutional network approach. Journal of Neural Engineering (2024)
8. Heinsfeld, A.S., Franco, A.R., Craddock, R.C., et al.: Identification of autism spectrum disorder using deep learning and the abide dataset. NeuroImage: Clinical (2018)
9. Hull, J.V., Dokovna, L.B., Jacokes, Z.J., et al.: Resting-state functional connectivity in autism spectrum disorders: a review. Frontiers in Psychiatry (2017)
10. Jeong, A.Y., Heo, D.W., Kang, E., et al.: Brainwavenet: Wavelet-based transformer for autism spectrum disorder diagnosis. In: Medical Image Computing and Computer-Assisted Intervention (2024)
11. Kan, X., Dai, W., Cui, H., et al.: Brain network transformer. Neural Information Processing Systems (2022)
12. Kim, B.H., Ye, J.C., Kim, J.J.: Learning dynamic graph representation of brain connectome with spatio-temporal attention. Neural Information Processing Systems (2021)
13. Kingma, D.P.: Adam: A method for stochastic optimization. International Conference on Learning Representations (2015)
14. Kipf, T.: Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations (2016)
15. Li, L., Wang, H., Zhang, W., et al.: Stg-mamba: Spatial-temporal graph learning via selective state space model. arXiv preprint arXiv:2403.12418 (2024)
16. Li, T., Le, W., Jankovic, J.: Linking the cerebellum to parkinson disease: an update. Nature Reviews Neurology (2023)

17. Li, X., Zhou, Y., Dvornek, N., et al.: Braingnn: Interpretable brain graph neural network for fmri analysis. Medical Image Analysis (2021)
18. Lin, H.Y., Tseng, W.Y.I., Lai, M.C., et al.: Altered resting-state frontoparietal control network in children with attention-deficit/hyperactivity disorder. Journal of the International Neuropsychological Society (2015)
19. Lin, T.Y., Goyal, P., Girshick, R., et al.: Focal loss for dense object detection. In: International Conference on Computer Vision (2017)
20. Marek, K., Jennings, D., Lasch, S., et al.: The parkinson progression marker initiative (ppmi). Progress in Neurobiology (2011)
21. Mills, K.L., Bathula, D., Dias, T.G.C., et al.: Altered cortico-striatal–thalamic connectivity in relation to spatial working memory capacity in children with adhd. Frontiers in Psychiatry (2012)
22. Paloyelis, Y., Mehta, M.A., Kuntsi, J., et al.: Functional mri in adhd: a systematic literature review. Expert Review of Neurotherapeutics (2007)
23. Pechlivanidou, G., Karampetakis, N.: Zero-order hold discretization of general state space systems with input delay. IMA Journal of Mathematical Control and Information (2022)
24. Peng, C., Huang, Y., Dong, Q., et al.: Biologically plausible brain graph transformer. International Conference on Learning Representations (2025)
25. Philip, R.C., Dauvermann, M.R., Whalley, H.C., et al.: A systematic review and meta-analysis of the fmri investigation of autism spectrum disorders. Neuroscience & Biobehavioral Reviews (2012)
26. Selvaraju, R.R., Cogswell, M., Das, A., et al.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: International Conference on Computer Vision (2017)
27. Sim, J., Jeon, S., Choi, I., et al.: Learning to approximate adaptive kernel convolution on graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence (2024)
28. Sim, J., Lee, M., Wu, G., Kim, W.H.: Multi-modal graph neural network with transformer-guided adaptive diffusion for preclinical alzheimer classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2024)
29. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., et al.: Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. Neuroimage (2002)
30. Veličković, P., Cucurull, G., Casanova, A., et al.: Graph attention networks. International Conference on Learning Representations (2017)
31. Wang, C., Tsepa, O., Ma, J., et al.: Graph-mamba: Towards long-range graph sequence modeling with selective state spaces. arXiv preprint arXiv:2402.00789 (2024)
32. Wei, Y., Abrol, A., Calhoun, V.D.: Hierarchical spatio-temporal state-space modeling for fmri analysis. In: International Conference on Research in Computational Molecular Biology (2025)
33. Wichmann, T., Bergman, H., DeLong, M.: The primate subthalamic nucleus. iii. changes in motor behavior and neuronal activity in the internal pallidum induced by subthalamic inactivation in the mptp model of parkinsonism. Journal of Neurophysiology (1994)
34. Wolters, A.F., van de Weijer, S.C., Leentjens, A.F., et al.: Resting-state fmri in parkinson's disease patients with cognitive impairment: a meta-analysis. Parkinsonism & Related Disorders (2019)

35. Wu, T., Long, X., Zang, Y., et al.: Regional homogeneity changes in patients with parkinson's disease. Human Brain Mapping (2009)
36. Xu, J., Yang, Y., Huang, D., et al.: Data-driven network neuroscience: On data collection and benchmark. Neural Information Processing Systems (2023)
37. Yang, L., Liu, Y., Zhuo, J., et al.: Do we really need message passing in brain network modeling? International Conference on Learning Representations (2025)
38. Yu, S., Jin, S., Li, M., et al.: Long-range brain graph transformer. Neural Information Processing Systems (2024)