MIXTURE OF WAVELETS ON SIMPLICIAL COMPLEX FOR ANALYSIS OF BRAIN CONNECTOME WITH NEURODEGENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Comprehensive analysis of brain networks is essential for understanding neurodegenerative disease progression and their early diagnosis. While various Graph Neural Networks (GNNs) have shown promise in identifying disease-related biomarkers, most existing approaches focus primarily on node-level features, overlooking crucial edge-level interactions. Furthermore, many spectral GNNs rely on fixed-bandwidth filtering, limiting their ability to capture diverse frequency components in graph-structured data. To address these limitations, we propose a novel framework that integrates two key components: 1) a Spectral Simplicial Wavelet Transform (SSWT) for jointly analyzing node and edge features, and 2) a Scale-aware eXpert on simplicial complex (ScaleX) for adaptive multi-scale filtering. Our approach utilizes richer spectral decomposition and dynamically selects the most relevant wavelet scales for each graph. Extensive experiments on benchmark brain network datasets demonstrate its effectiveness in improving classification and interpretability, highlighting its potential for facilitating the analysis of neurodegenerative diseases. *The code will be released upon publication.*

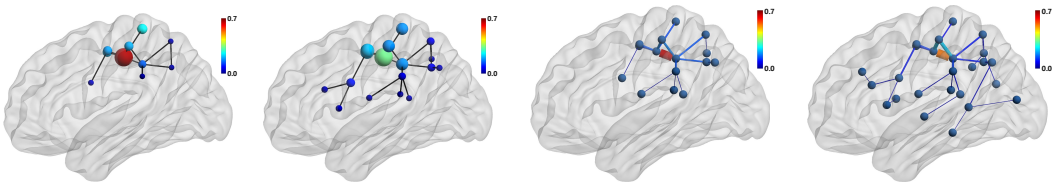
1 INTRODUCTION

Brain connectome provides a comprehensive map of the neural connections by connecting different anatomical regions of interest (ROIs) in the brain (Bullmore et al., 2011). It offers an informative feature for understanding neurodegenerative diseases such as Alzheimer's Disease (AD) and Parkinson's Disease (PD) as they progressively impair brain function by disrupting neural connections (Pievani et al., 2014). The connectome defines a brain network represented as a graph, where nodes are ROIs and edges define connection strengths derived from neuroimaging modalities such as Diffusion Weighted Image (DWI) and functional Magnetic Resonance Imaging (fMRI). While analyzing these graphs helps identify disease-specific variations and supports early diagnosis and monitoring (Tijms et al., 2013), it often remains challenging due to the complex structure of brain networks and the inherent variability across individuals.

Recently, Graph Neural Network (GNN) and its variants have gained significant attention in brain network analysis (Kan et al., 2022; Yu et al., 2024). While many of them operate in the naive graph space (Li et al., 2021), some utilize the spectral space of brain networks to address the heterogeneity of individual graphs, offering a powerful framework for modeling intricate topological structures and capturing underlying spectral properties (He et al., 2020). They leverage graph Fourier transform and filtering in the graph Fourier space to analyze signals on the graph, enabling the identification of critical regions associated with disease progression. The choice of filters and their scale (i.e., bandwidths) selectively captures global and localized features across different frequency components, highlighting disease-related alterations. Such approaches have shown promise in identifying early biomarkers of neurodegenerative diseases, even under noise and inter-individual variability.

Most GNNs, both the spatial and spectral approaches, predominantly focus on node-level analyses, often treating edge features as auxiliary information within the adjacency matrix (Kipf & Welling, 2017; Gasteiger et al., 2019). While these methods are effective for widely used graph benchmarks where edges represent simple categorical properties such as bond types or binary connectivity (Morris et al., 2020; Hu et al., 2020), their applicability is limited in brain networks. Edges in brain networks represent rich, continuous-valued connectomic features that capture the integrity of con-

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.



(a) Node-localized wavelet basis, $s = 1$ (b) Node-localized wavelet basis, $s = 4$ (c) Edge-localized wavelet basis, $s = 1$ (d) Edge-localized wavelet basis, $s = 4$

Figure 1: Visualization of diffusion filters (i.e., $\mathcal{K}_s(\lambda) = e^{-s\lambda}$) localized at a node (a and b) and an edge (c and d) in different scales. Smaller scales cover a narrow region, while larger scales enable broader diffusion, smoothing features across large neighborhoods (low-pass filtering). Node size and color indicate node activation, and edge thickness and color represent edge activation.

nections, providing critical insights into disease progression. While recent studies have explored various approaches to incorporate edge-level signals, including line graph-based formulations and Hodge Laplacian-based methods (Cai et al., 2021; Roddenberry et al., 2022), these approaches have largely focused on single-order processing, with little effort made to develop a joint spectral framework for node- and edge-level signals.

Moreover, most existing spectral GNNs rely on fixed-bandwidth filtering, limiting their ability to capture diverse frequency components in graph-structured data. While polynomial approximations (Defferrard et al., 2016) and kernel-based methods (Xu et al., 2019) offer some flexibility, they typically apply uniform spectral filters across all nodes. This rigidity prevents models from adapting to regions that require different frequency responses—such as low-pass filtering for capturing global structures and band-pass filtering for preserving local variations—ultimately leading to suboptimal representations. More importantly, these methods lack mechanisms to dynamically adjust the filtering bandwidth and limits adaptability to diverse spectral patterns.

To address these limitations, we propose 1) **S**pectral **S**implicial **W**avelet **T**ransform (SSWT) to derive multi-resolution node and edge features, and 2) **Scale**-aware e**X**pert on simplicial complex (ScaleX) for adaptive multi-scale analysis. Leveraging the r-Laplacian (i.e., Hodge Laplacian), SSWT provides a richer multi-scale representation of brain networks explicitly on nodes and edges, beyond traditional node-centric approaches. SSWT decomposes a graph into local-to-global connectivity patterns of both nodes and edges at different resolutions with wavelets. Fig. 1 visualizes the spectral filters defined over the r-Laplacians, where (a) and (b) show localized filters on nodes (i.e., $r = 0$), while (c) and (d) show localized filters on edges (i.e., $r = 1$). Furthermore, to handle spectral heterogeneity in graph analysis, we introduce ScaleX, which sparsely selects spectral experts based on the compatibility between individual node and edge signals and learnable wavelet scales. By prioritizing informative scales for individual graph elements, ScaleX adapts spectral filtering to heterogeneous local spectral characteristics within each brain network and enhances generalizability across diverse neuroimaging datasets.

To this end, we introduce an integrated framework with the following **contributions**: **1)** We extend the Spectral Graph Wavelet Transform (SGWT) to a learnable Spectral Simplicial Wavelet Transform (SSWT) via Hodge Laplacian, enabling adaptive spectral analysis of both node and edge features in brain networks. **2)** We propose a Scale-aware eXpert on simplicial complex (ScaleX) that adaptively selects wavelet scales for individual nodes and edges through sparse signal-scale routing, capturing diverse spectral resolutions for comprehensive graph representation. **3)** We provide theoretical grounding for our framework, demonstrating that node and edge spectral representations are aligned through a shared incidence-based spectrum with orientation-equivariant edge-level filtering. Extensive experiments on ADNI and PPMI datasets show that our model achieves state-of-the-art performance and provides clinically interpretable results in brain regional and connectomic analysis.

2 Related Work

Spectral Graph Neural Networks. Spectral GNNs analyze graph-structured data by leveraging spectral representations grounded in graph signal processing (Shuman et al., 2013). Early spectral GNNs relied on fixed-bandwidth filters to capture graph characteristics. ChebNet (Defferrard et al., 2016) introduced Chebyshev polynomial filters to approximate graph convolutions, while GCN (Kipf & Welling, 2017) further simplified this approach via first-order approximation. Diffusion-based methods, such as GraphHeat (Xu et al., 2019) and GDC (Gasteiger et al., 2019),

leverage the heat kernel to enhance low-pass properties in graph structures. Recent works, such as Exact (Choi et al., 2022) and AGT (Cho et al., 2024), introduced adaptive spectral filters with trainable parameters, enabling dynamic adjustment of frequency bandwidths based on the graph structure. Furthermore, Hodgelet (Roddenberry et al., 2022) proposed Hodge Laplacian-based fixed spectral wavelet dictionaries for edge flows on graphs focusing on signal representation and sparse coding. Unlike these methods, our work introduces multi-order spectral graph wavelets as learnable, multi-scale filters that are jointly applied to node and edge signals.

Mixture of Experts. The Mixture of Experts (MoE) framework enhances model performance by dividing a complex task into smaller, specialized sub-tasks (Jacobs et al., 1991; Jordan et al., 1994). Early MoE models used traditional machine learning techniques (Jordan et al., 1996; Chen et al., 1999), while modern variants incorporate sparse activation, where only a subset of experts is utilized for each input, significantly improving efficiency (Shazeer et al., 2017; Roller et al., 2021). It has been particularly successful in large-scale applications such as natural language processing (NLP), enabling the training of extremely large models (Fedus et al., 2022). Beyond NLP, MoE has also been adopted in computer vision (Riquelme et al., 2021; Zamfir et al., 2024), graph (Wang et al., 2023; Guo et al., 2025), and multi-modal learning (Shi et al., 2019; Mustafa et al., 2022).

3 PRELIMINARY

Hodge Laplacians for Graphs. A simplicial complex is a structure that represents data as collections of simplices, which are fundamental building blocks such as points (0-simplex), lines (1-simplex), triangles (2-simplex), and their higher-dimensional counterparts. For any simplex included in a simplicial complex, all its faces (lower-dimensional simplices) must also be part of the complex. A graph can be viewed as a 1-dimensional simplicial complex, consisting only of 0-simplices (nodes) and 1-simplices (edges) (Schaub et al., 2020; Anand & Chung, 2023).

To formally describe relationships between r-simplices and $(r-1)$-simplices, we use the boundary operator $\partial_r : C_r \to C_{r-1}$, where C_r denotes the space of r-chains. An r-chain is a formal linear combination of r-simplices, written as $c = \sum_i \alpha_i \sigma_i^r$, where σ_i^r is an r-simplex and α_i is a coefficient. To define boundaries consistently, simplices are assumed to be oriented, ensuring that shared faces cancel properly, i.e., $\partial_{r-1} \circ \partial_r = 0$. For an oriented r-simplex $\sigma^r = [v_0, v_1, \ldots, v_r]$, its boundary is given by $\partial_r(\sigma^r) = \sum_{i=0}^{r} (-1)^i [v_0, v_1, \ldots, \hat{v}_i, \ldots, v_r]$, where $\hat{v}_i$ denotes the omission of the i-th vertex. The boundary operator can be represented as a boundary matrix B_r, which relates r-simplices to $(r-1)$-simplices (e.g., edge-to-node connection). The elements of B_r are defined as:

$$B_r(i,j) = \begin{cases} 1 & \text{if } \sigma_i^{r-1} \subset \sigma_j^r \text{ and orientations align,} \\ -1 & \text{if } \sigma_i^{r-1} \subset \sigma_j^r \text{ and orientations differ,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Based on B_r, the Hodge Laplacian for r-simplices is defined as $\mathcal{L}_r = B_r^T B_r + B_{r+1} B_{r+1}^T$. This extends the graph Laplacian to higher-order structures, enabling analysis of multi-dimensional interactions (Dakurah et al., 2022). For graphs with 0-simplices (nodes) and 1-simplices (edges), $B_2 = 0$, yielding $\mathcal{L}_0 = B_1 B_1^T$ and $\mathcal{L}_1 = B_1^T B_1$, where B_1 is the node-to-edge incidence matrix. Here, $\mathcal{L}_0$ is the graph Laplacian capturing node-level connectivity, while $\mathcal{L}_1$ captures edge-level relationships.

Wavelet Transform in Signal Processing. The wavelet transform analyzes signals at multiple resolutions, providing simultaneous localization in both time (or space) and frequency (Mallat, 1989). A mother wavelet ψ (with a band-pass characteristic), defined over a domain variable x (e.g., time or spatial location), with scale s and translation τ is given as:

$$\psi_{s,\tau}(x) = \frac{1}{s} \psi \left(\frac{x - \tau}{s} \right), \tag{2}$$

where s controls the dilation of the wavelet $\psi_{s,\tau}$. Small s corresponds to high frequencies (capturing fine details) and large s corresponds to low frequencies (capturing global trends). τ shifts the wavelet in space, enabling localized signal analysis. For a signal $f(x)$, its wavelet coefficient is defined as:

$$W_f(s,\tau) = \langle f, \psi_{s,\tau} \rangle = \frac{1}{s} \int f(x) \psi^* \left(\frac{x - \tau}{s} \right) dx, \tag{3}$$

where * denotes complex conjugate. It is invertible, allowing the original signal to be reconstructed:

$$f(x) = \frac{1}{C_\psi} \int \int W_f(s,\tau) \psi_{s,\tau}(x) d\tau \, ds, \quad C_\psi = \int \frac{|\Psi(j\omega)|^2}{|\omega|} d\omega, \tag{4}$$

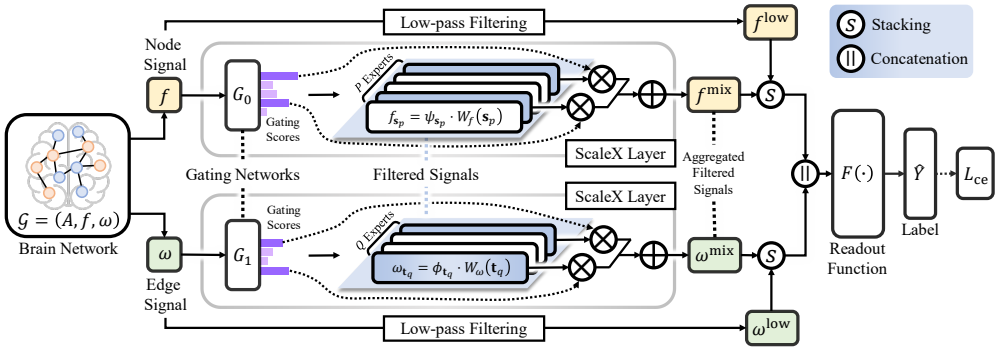


Figure 2: **Overall framework of our proposed method.** Given a brain network $\mathcal{G}$ with node signal f and edge signal ω, the node and edge routing operations in Eqs. (9)–(10), denoted by G_0 and G_1 in the figure, select scale experts. The selected wavelet bases $\psi_{\mathbf{s}_p}$ and $\phi_{\mathbf{t}_q}$ generate scale-filtered signals, which are aggregated using the routing weights (gating scores), combined with f^{low} and ω^{low}, and passed to the readout function $F(\cdot)$. For clarity, the use of $\bar{\omega}$ for edge routing, orientation alignment, and $\mathcal{L}_{\text{bal}}$ (Sec. 4.4) are omitted.

where $C_\psi < \infty$ is the admissibility constant with Ψ, i.e., the Fourier transform of the wavelet function ψ in the frequency space ω. An additional scaling function φ is often introduced as a low-pass filter to capture low-frequency components that mother wavelets cannot handle.

4 Method

Given a brain network with brain ROIs as nodes, their connections as edges, and associated node and edge signals, we propose a learnable higher-order extension of Spectral Graph Wavelet Transform (SGWT) (Hammond et al., 2011), as illustrated in Fig. 2. In Sec. 4.1, we extend SGWT with Hodge Laplacians to decompose node and edge signals at specific scales. Sec. 4.2 introduces adaptive wavelet bases to learn optimal scales for each node and edge, while Sec. 4.3 presents a Scale-aware eXpert on simplicial complex (ScaleX) that dynamically selects relevant scales for multi-resolution analysis. Finally, Sec. 4.4 presents the training objective with routing balance regularization for balanced expert utilization. While this work focuses on node- and edge-level interactions, extension to higher-order simplicial complexes is discussed in Sec. B.

4.1 Spectral Simplicial Wavelet Transform

Let $\mathcal{G} = (A, f, \bar{\omega})$ be an undirected graph, where $A \in \mathbb{R}^{N\times N}$ is an adjacency matrix that defines the connectivity between N nodes, $f \in \mathbb{R}^{N\times d_f}$ represents d_f-dimensional node signals, and $\bar{\omega} \in \mathbb{R}^{M\times d_\omega}$ denotes d_ω-dimensional edge signals for M edges. Since the edges of an undirected graph have no intrinsic orientation, we assign an arbitrary orientation to construct the boundary matrix $B_1 \in \mathbb{R}^{N\times M}$. Let $S = \text{diag}(\pm 1) \in \mathbb{R}^{M\times M}$ encode this orientation relative to a fixed reference orientation, and define the oriented edge signal as $\omega = S\bar{\omega}$. Using B_1, the Hodge Laplacians $\mathcal{L}_0$ and $\mathcal{L}_1$ are defined as in Sec. 3. Both $\mathcal{L}_0$ and $\mathcal{L}_1$ are real and positive semi-definite and can be decomposed into $\mathcal{L}_0 = U_0\Lambda_0 U_0^T$ and $\mathcal{L}_1 = U_1\Lambda_1 U_1^T$, where $U_0 = [u_{0,1}, \ldots, u_{0,N}]$ and $U_1 = [u_{1,1}, \ldots, u_{1,M}]$ are orthonormal matrices, and $\Lambda_0 = \text{diag}(\lambda_{0,1}, \ldots, \lambda_{0,N})$ and $\Lambda_1 = \text{diag}(\lambda_{1,1}, \ldots, \lambda_{1,M})$ are diagonal matrices containing the non-negative eigenvalues.

Proposition 1. *Let $B_1 = U\Sigma V^T$ be the singular value decomposition (SVD) of the incidence matrix. Then, the eigenvectors of Hodge Laplacians $\mathcal{L}_0$ and $\mathcal{L}_1$ are given by U and V, respectively (i.e., $U_0 = U, U_1 = V$), and excluding the zero eigenvalues, both Laplacians share the same non-zero eigenvalues $\lambda_i = \sigma_i^2$, where σ_i are the non-zero singular values from Σ.*

This proposition implies that the spectral analysis of nodes and edges can be unified under a single singular value spectrum, avoiding separate eigen-decompositions of $\mathcal{L}_0$ and $\mathcal{L}_1$. (Proof is in Sec. A)

To analyze the node signal f and the oriented edge signal ω at multiple scales, we extend the SGWT to Spectral Simplicial Wavelet Transform (SSWT). SSWT enables the decomposition of multi-order graph signals into multi-scale components. The wavelet bases for nodes and edges are defined as:

$$\psi_s = U\mathcal{K}(s\Lambda_0)U^T, \quad \phi_t = V\mathcal{K}(t\Lambda_1)V^T, \tag{5}$$

where $\mathcal{K}$ is a kernel function and $s, t \in \mathbb{R}$ are scale parameters. These basis functions (also visualized in Fig. 1) are localized in the graph space, capturing features specific to individual nodes (δ_n) or

edges (δ_m) at a fixed resolution. f and ω are projected onto the spectral domain using these bases, yielding wavelet coefficients $W_f(s) = \psi_s \cdot f$ and $W_\omega(t) = \phi_t \cdot \omega$. Under the admissibility condition, the signals can be reconstructed via the Inverse Simplicial Wavelet Transform (ISWT):

$$f = \frac{1}{C_\mathcal{K}} \int_0^\infty \psi_s \cdot W_f(s) \frac{ds}{s}, \quad \omega = \frac{1}{C_\mathcal{K}} \int_0^\infty \phi_t \cdot W_\omega(t) \frac{dt}{t}, \tag{6}$$

where $C_\mathcal{K} = \int_0^\infty \frac{\mathcal{K}^2(\lambda)}{\lambda} d\lambda < \infty$. Eq. (6) represents the superposition of multi-resolution representations of f and ω over scales $s, t \in [0, \infty)$. Each component at scales s and t is given as:

$$f_s = \psi_s \cdot W_f(s) = U\mathcal{K}^2(s\Lambda_0)U^T f, \quad \omega_t = \phi_t \cdot W_\omega(t) = V\mathcal{K}^2(t\Lambda_1)V^T\omega. \tag{7}$$

This enables the extraction of information corresponding to the desired resolution of signals, allowing for detailed multi-resolution analysis in the spatial domain.

Orientation Equivariance. For an undirected graph, constructing B_1 requires assigning an orientation to each edge. While $\mathcal{L}_0$ is invariant to this choice, $\mathcal{L}_1$ and the oriented edge signal ω depend on the assigned edge orientations. This dependency challenges the consistency of graph analysis under arbitrary orientation choices. To address this, we show that SSWT satisfies *orientation equivariance*, a property studied in simplicial neural networks (Schaub et al., 2020; Yang et al., 2022). Specifically, reorienting a subset of edges results in sign flips in the filtered edge representations.

Proposition 2. *Consider a boundary matrix $B_1 \in \mathbb{R}^{N\times M}$ with arbitrary orientations and edge signal ω of a graph $\mathcal{G}$. Let $J \in \mathbb{R}^{M\times M}$ be an arbitrary diagonal matrix encoding changes in edge orientation, where $J[i,i] = -1$ if the orientation of the i-th edge is flipped, and $J[i,i] = 1$ otherwise. Define the reoriented graph $\mathcal{G}'$, with $\omega' = J\omega$ and $B_1' = B_1 J$. Then, the filtered edge signal ω_t' on $\mathcal{G}'$ satisfies $\omega_t' = J\omega_t$, demonstrating that the ISWT is equivariant to edge orientation.*

Proposition 2 shows that flipping the orientation of certain edges changes the sign of the corresponding components of ω_t, while preserving their magnitudes as in the proof in Sec. A. This ensures consistent edge-level spectral filtering under arbitrary orientation choices in undirected settings.

4.2 Adaptive Scaling for Nodes and Edges

Conventional graph convolution aggregates information uniformly across nodes with respect to the number of hidden layers, which often results in over- or under-smoothing, particularly when there are significant local variations. To address this, we adopt an adaptive scaling approach for both nodes and edges, allowing them to aggregate information optimally based on their local graph structure. We define node-wise and edge-wise wavelet bases using a trainable set of scaling parameters, $\mathbf{s} = \{s_i\}_{i=1}^N$ associated with nodes $\{n_i\}_{i=1}^N$ and $\mathbf{t} = \{t_i\}_{i=1}^M$ associated with edges $\{e_i\}_{i=1}^M$. The adaptive wavelet bases ψ_{s_i} at n_j localized at n_i and ϕ_{t_i} at e_j localized at e_i are formulated as:

$$\psi_{s_i,n_i}(n_j) = \sum_{\ell=1}^N \mathcal{K}(s_i\lambda_{0,\ell}) u_{0,\ell}^*(n_i) u_{0,\ell}(n_j), \quad \phi_{t_i,e_i}(e_j) = \sum_{\ell=1}^M \mathcal{K}(t_i\lambda_{1,\ell}) u_{1,\ell}^*(e_i) u_{1,\ell}(e_j), \tag{8}$$

By adjusting the scales s_i and t_i, this approach enhances or suppresses specific details in the spatial domain as needed, enabling a more nuanced analysis of graph signals.

4.3 Scale-aware Expert on Simplicial Complex

The key to effectively analyzing brain networks lies in understanding their spectral multi-scale characteristics, which play a crucial role in capturing both global and local graph patterns. We propose ScaleX, a framework that dynamically combines spectral information across multiple scales. We define multiple learnable sets of such adaptive wavelet scales for nodes and edges, treating each set as an expert. ScaleX employs sparse routing to adaptively select relevant scale experts based on their compatibility with individual node and edge signals. The filtering results at the selected scales are combined into multi-resolution graph features, capturing diverse scale-specific characteristics.

Signal-scale routing. To adapt spectral scales to individual nodes and edges, we introduce a routing mechanism that learns the compatibility between graph signals and spectral scale sets. For each node i and node expert p, we compute the signal-scale compatibility score as

$$\ell_{i,p}^{(0)} = \frac{\varphi_0(f_i)^\top \chi_0(\mathbf{s}_p)}{\sqrt{d_{route}}}, \qquad \boldsymbol{\alpha}_i^{(0)} = \mathrm{Softmax}\left(\mathrm{TopK}(\boldsymbol{\ell}_i^{(0)}, k)\right), \tag{9}$$

where φ_0 and χ_0 project the node signal and its learnable spectral scale set into a shared d_{route}-dimensional routing space, respectively. The TopK operator retains the k largest compatibility

scores and sets the others to $-\infty$. Similarly, for each edge e and edge expert q, we compute

$$\ell_{e,q}^{(1)} = \frac{\varphi_1(\bar{\omega}_e)^\top \chi_1(\mathbf{t}_q)}{\sqrt{d_{route}}}, \qquad \boldsymbol{\alpha}_e^{(1)} = \mathrm{Softmax}\left(\mathrm{TopK}(\boldsymbol{\ell}_e^{(1)}, k)\right), \tag{10}$$

where φ_1 and χ_1 project the orientation-agnostic edge signal and its learnable spectral scale set into the shared routing space. Using $\bar{\omega}_e$ for routing separates the intrinsic edge signal from the arbitrary orientation introduced for spectral filtering. Consequently, edge-scale selection remains invariant to edge reorientation, while the corresponding SSWT features preserve orientation equivariance.

ScaleX layer. The ScaleX layer aggregates multi-scale spectral features for both nodes and edges using the signal-scale routing weights defined in Eqs. (9) and (10). With P scale sets for nodes and Q scale sets for edges, the aggregated representations for node i and edge e are computed as

$$f_i^{\mathrm{mix}} = \sigma\left(\sum_{p=1}^{P} \alpha_{i,p}^{(0)} f_{\mathbf{s}_p,i}\theta_0\right), \qquad \omega_e^{\mathrm{mix}} = \sigma\left(\sum_{q=1}^{Q} \alpha_{e,q}^{(1)} \omega_{\mathbf{t}_q,e}\theta_1\right), \tag{11}$$

where $f_{\mathbf{s}_p,i}$ and $\omega_{\mathbf{t}_q,e}$ denote the node and edge features from the adaptive wavelet bases in Eq. (8) with scale sets $\mathbf{s}_p$ and $\mathbf{t}_q$, respectively. Here, $\sigma(\cdot) = \tanh(\cdot)$ denotes the activation function, whose odd symmetry preserves the equivariance of the aggregated edge features. The matrices $\theta_0 \in \mathbb{R}^{d_f \times h}$ and $\theta_1 \in \mathbb{R}^{d_\omega \times h}$ are learnable weight matrices shared across all scales in the h-dimensional space. Since $\alpha_{e,q}^{(1)}$ is invariant to edge reorientation and $\omega_{\mathbf{t}_q,e}$ is orientation-equivariant by Proposition 2, ω_e^{mix} adaptively integrates scale-specific features while preserving orientation equivariance.

Choice of Kernels. A single low-pass and multiple band-pass wavelet kernels are utilized to decompose the input signals f and ω. For low-pass filtering, we adopt the kernel $\mathcal{K}_s^{\mathrm{low}}(\lambda) = e^{-s\lambda}$ as a scaling function, which smooths local variations and emphasizes global structures by aggregating information over a broad range of nodes and edges. To capture band-pass characteristics, the kernel $\mathcal{K}_s(\lambda) = s\lambda e^{-s\lambda}$ is employed, emphasizing intermediate frequencies and highlighting localized features in the graph. A detailed rationale for these kernel selections is provided in Sec. C.

In low-pass filtering, the scales $\mathbf{s}^{\mathrm{low}}$ for nodes and $\mathbf{t}^{\mathrm{low}}$ for edges are consistently learned and applied across the graph. The filtered representations are computed as $f^{\mathrm{low}} = \sigma(f_{\mathbf{s}^{\mathrm{low}}}\theta_0)$ and $\omega^{\mathrm{low}} = \sigma(\omega_{\mathbf{t}^{\mathrm{low}}}\theta_1)$, where $f_{\mathbf{s}^{\mathrm{low}}} \in \mathbb{R}^{N \times d_f}$ and $\omega_{\mathbf{t}^{\mathrm{low}}} \in \mathbb{R}^{M \times d_\omega}$ represent node and edge features filtered using the low-pass kernel according to Eq. (7). For band-pass filtering, multiple scales are adaptively combined using ScaleX layer, as defined in Eq. (11). Finally, these outputs are stacked to form the final node and edge features as $\tilde{f} = [f^{\mathrm{low}}, f^{\mathrm{mix}}] \in \mathbb{R}^{N \times 2h}$ and $\tilde{\omega} = [\omega^{\mathrm{low}}, \omega^{\mathrm{mix}}] \in \mathbb{R}^{M \times 2h}$. These embeddings balance global and localized information to improve downstream tasks.

4.4 Training Objective with Routing Balance

Routing balance loss. To prevent the signal-scale routing from collapsing to a small subset of scale experts, we introduce a balance regularization over the routing weights. For a mini-batch, let $\mathcal{V}_\mathcal{B}$ and $\mathcal{E}_\mathcal{B}$ denote the sets of nodes and edges, respectively. We compute the average routing weights for each node and edge scale expert as:

$$\bar{\alpha}_p^{(0)} = \frac{1}{|\mathcal{V}_\mathcal{B}|} \sum_{i \in \mathcal{V}_\mathcal{B}} \alpha_{i,p}^{(0)}, \qquad \bar{\alpha}_q^{(1)} = \frac{1}{|\mathcal{E}_\mathcal{B}|} \sum_{e \in \mathcal{E}_\mathcal{B}} \alpha_{e,q}^{(1)}. \tag{12}$$

We then encourage balanced utilization of scale experts across the mini-batch by minimizing

$$L_{\mathrm{bal}} = \sum_{p=1}^{P} \left(\bar{\alpha}_p^{(0)} - \frac{1}{P}\right)^2 + \sum_{q=1}^{Q} \left(\bar{\alpha}_q^{(1)} - \frac{1}{Q}\right)^2. \tag{13}$$

This regularization prevents routing collapse at the batch level while allowing individual nodes and edges to maintain distinct and sparse scale preferences.

Training Objective. The goal of the model is to predict the true graph label Y as $\hat{Y}$. Before graph-level prediction, we align the orientation-equivariant edge embedding $\tilde{\omega}$ with the reference orientation as $\omega^{\mathrm{align}} = S\tilde{\omega}$. Under any edge reorientation J, we have $S' = JS$ and $\tilde{\omega}' = J\tilde{\omega}$, yielding $S'\tilde{\omega}' = S\tilde{\omega}$. The combined graph embedding $[\tilde{f}; \omega^{\mathrm{align}}] \in \mathbb{R}^{(N+M)\times 2h}$ is then passed through a readout function $F(\cdot)$ to obtain $\hat{Y}$. Using cross-entropy, the prediction loss is defined as:

$$L_{\mathrm{ce}} = -\sum_{b=1}^{\mathcal{B}} \sum_{c \in C} Y_{bc} \log(\hat{Y}_{bc}), \tag{14}$$

Table 1: Classification performance on the ADNI and PPMI datasets evaluated using 5-fold cross-validation with 10 random seeds. The best results are shown in **bold**, and the second-best results are underlined.

Model	ADNI-CT			ADNI-FDG			PPMI		
	Accuracy (%)	Precision (%)	Recall (%)	Accuracy (%)	Precision (%)	Recall (%)	Accuracy (%)	Precision (%)	Recall (%)
SVM (Linear)	82.5 ± 2.0	86.2 ± 3.3	82.2 ± 2.3	85.3 ± 2.1	85.7 ± 2.7	86.9 ± 2.1	60.5 ± 10.1	53.1 ± 6.6	51.9 ± 8.2
MLP (2-layers)	78.8 ± 2.2	79.2 ± 3.6	79.9 ± 2.6	87.5 ± 1.6	88.2 ± 2.4	88.1 ± 1.4	68.9 ± 3.5	59.1 ± 4.4	59.7 ± 8.6
GCN (Kipf & Welling, 2017)	59.9 ± 1.8	59.8 ± 2.5	57.4 ± 2.3	68.8 ± 2.0	67.7 ± 2.8	69.7 ± 2.5	78.8 ± 2.1	74.7 ± 5.2	74.5 ± 4.3
GAT (Veličković et al., 2018)	64.2 ± 5.5	62.7 ± 6.7	66.8 ± 4.6	69.2 ± 7.1	67.0 ± 10.6	73.6 ± 3.7	81.2 ± 2.4	75.0 ± 6.9	77.5 ± 5.5
GDC (Gasteiger et al., 2019)	77.1 ± 4.3	76.9 ± 5.0	78.5 ± 4.4	86.2 ± 3.2	86.7 ± 3.3	87.0 ± 2.9	73.0 ± 0.7	69.9 ± 3.1	64.9 ± 9.4
GraphHeat (Xu et al., 2019)	70.9 ± 3.2	70.3 ± 3.0	71.8 ± 2.6	77.0 ± 2.4	77.5 ± 3.5	77.3 ± 1.0	79.1 ± 2.0	70.7 ± 4.7	75.4 ± 3.0
ADC (Zhao et al., 2021)	82.1 ± 2.4	77.6 ± 1.9	72.8 ± 6.7	88.6 ± 2.8	87.9 ± 6.2	87.2 ± 5.3	78.8 ± 2.3	70.5 ± 8.0	72.0 ± 5.5
BrainGNN (Li et al., 2021)	43.5 ± 1.6	35.5 ± 0.8	37.4 ± 2.1	68.9 ± 2.4	66.0 ± 0.4	64.5 ± 13.3	69.6 ± 5.1	61.2 ± 7.9	65.0 ± 5.0
BrainNetTF (Kan et al., 2022)	87.8 ± 3.9	86.8 ± 16.7	86.2 ± 9.0	87.1 ± 5.3	86.1 ± 16.5	85.4 ± 15.6	71.3 ± 4.1	64.2 ± 10.5	66.1 ± 6.3
Exact (Choi et al., 2022)	86.2 ± 2.0	86.6 ± 1.7	86.7 ± 2.3	90.2 ± 2.7	90.7 ± 2.8	90.7 ± 2.8	79.5 ± 2.4	71.9 ± 5.4	71.4 ± 9.0
SCNN (Yang et al., 2022)	87.0 ± 1.1	87.2 ± 2.1	87.9 ± 1.3	92.3 ± 1.8	92.5 ± 2.2	93.1 ± 1.8	79.8 ± 0.7	75.7 ± 4.7	75.4 ± 5.4
SaNN (Gurugubelli & Chepuri, 2024)	84.4 ± 2.1	85.0 ± 3.0	85.7 ± 1.5	92.1 ± 1.8	92.2 ± 2.7	93.2 ± 1.4	77.4 ± 2.9	68.4 ± 6.4	69.5 ± 15.5
ALTER (Yu et al., 2024)	84.4 ± 2.2	83.5 ± 10.9	82.7 ± 11.2	85.1 ± 4.1	84.2 ± 12.5	83.6 ± 14.4	82.5 ± 6.3	77.8 ± 6.6	75.6 ± 4.7
ContrastPool (Xu et al., 2024)	63.3 ± 4.1	61.7 ± 2.9	64.0 ± 4.5	75.3 ± 1.3	74.7 ± 1.3	76.2 ± 1.9	71.3 ± 5.2	61.4 ± 5.9	64.9 ± 11.9
AGT (Cho et al., 2024)	90.3 ± 1.8	91.3 ± 2.4	89.6 ± 2.3	94.8 ± 1.1	94.3 ± 1.5	95.3 ± 1.4	83.6 ± 3.8	78.6 ± 4.4	79.8 ± 3.5
SSWT w/o ScaleX (ours)	92.5 ± 1.1	92.0 ± 1.1	93.4 ± 1.0	92.8 ± 0.7	93.3 ± 1.3	93.1 ± 0.9	84.6 ± 2.7	80.4 ± 5.5	80.2 ± 2.2
SSWT w/ ScaleX (ours)	**94.9 ± 1.3**	**95.4 ± 0.8**	**94.4 ± 1.7**	**96.4 ± 0.9**	**96.5 ± 1.1**	**97.1 ± 1.1**	**87.4 ± 2.1**	**82.1 ± 3.4**	**80.7 ± 1.6**

where Y_{bc} represents the true class label for the b-th sample and the c-th class, and $\hat{Y}_{bc}$ denotes the corresponding predicted probability. The overall objective is $L = L_{\text{ce}} + \lambda_{\text{bal}} L_{\text{bal}}$, where λ_{bal} controls the contribution of the routing balance regularization.

5 Experiment

5.1 Dataset and Experiment Settings

Two brain network datasets are used: 1) the Alzheimer's Disease Neuroimaging Initiative (ADNI) and 2) the Parkinson's Progression Markers Initiative (PPMI). Diffusion MRI in ADNI was used to construct structural connectomes, as they capture anatomical degeneration (Vemuri & Jack, 2010)—such as white matter disruption and cortical atrophy, and AD pathology is known to propagate along these structural pathways (Raj et al., 2012). To further assess the generalizability of our model, we also evaluate it on the functional connectivity from PPMI. Detailed demographics and graph construction are given in Sec. D. We further include results on the Autism Brain Imaging Data Exchange (ABIDE) (Heinsfeld et al., 2018) dataset in Sec. F.1.

ADNI Dataset. The ADNI study (Mueller et al., 2005) is a public repository for AD research that provides multi-modal imaging and biomarkers at various stages of cognitive decline. Structural brain networks are constructed using probabilistic tractography on DWI, with 160 brain regions from Destrieux atlas (Destrieux et al., 2010) as nodes, and white-matter fiber connections weighted by tract counts as edges. Node features include region-wise cortical thickness (CT) from MRI and Standardized Uptake Value Ratios (SUVRs) from FDG-PET. Subjects are grouped into five categories: Cognitively Normal (CN), Significant Memory Concern (SMC), Early Mild Cognitive Impairment (EMCI), Late MCI (LMCI), and AD.

PPMI Dataset. PPMI (Marek et al., 2011) provides imaging and biomarkers for PD progression. Functional brain networks are derived from fMRI on 116 regions based on the AAL atlas (Tzourio-Mazoyer et al., 2002), with Blood-Oxygen-Level-Dependent (BOLD) signals as node features and functional connectivity (correlation) between regions as edge features (Xu et al., 2023). The dataset contains three diagnostic groups: CN, Prodromal, and PD.

Implementation. All models were implemented in PyTorch and trained on a single NVIDIA RTX 6000 Ada Generation GPU using the Adam optimizer with early stopping for up to 5000 epochs. Hyperparameters were tuned via grid search: hidden units in $\{2, 4, 6, 8, 16, 32, 64\}$, dropout rates in $\{0.2, 0.4, 0.6, 0.8\}$, and learning rates in $\{0.1, 0.01, 0.001, 0.0001\}$. We further provide an analysis of runtime and memory usage in Sec. F. Additional implementation details, including expert configurations and hyperparameter settings for both our model and baselines, are provided in Sec. E.

Evaluation. We compared our method against a wide range of baselines, including recent state-of-the-art models. These include conventional classifiers (Linear SVM, 2-layer MLP); spatial GNNs (GCN (Kipf & Welling, 2017), GAT (Veličković et al., 2018)); spectral and diffusion-based GNNs (GDC (Gasteiger et al., 2019), GraphHeat (Xu et al., 2019), ADC (Zhao et al., 2021), Exact (Choi et al., 2022), AGT (Cho et al., 2024)); simplicial complex-based models (SCNN (Yang et al., 2022), SaNN (Gurugubelli & Chepuri, 2024)); and brain-specific models (BrainGNN (Li et al., 2021), BrainNetTF (Kan et al., 2022), ALTER (Yu et al., 2024), ContrastPool (Xu et al., 2024)). All baselines were implemented and tuned under comparable settings to ensure fair evaluation. (Details

Table 2: Grad-CAM outcomes of the top-10 ROIs and connectomes with the highest activation (Act.) for AD/PD from the ADNI-CT (left) / PPMI (right) results. The indices given in Destrieux and AAL atlas.

ADNI-CT						PPMI					
Idx	ROI (Node)	Act.	Idx1-Idx2	ROI1-ROI2 (Edge)	Act.	Idx	ROI (Node)	Act.	Idx1-Idx2	ROI1-ROI2 (Edge)	Act.
158	R-Thalamus_Proper	0.977	153-154	L-Putamen - L-Pallidum	0.750	74	Putamen_R	0.717	74-78	Putamen_R - Thalamus_R	0.887
156	R-Caudate	0.867	159-160	R-Putamen - R-Pallidum	0.715	77	Thalamus_L	0.688	73-77	Putamen_L - Thalamus_L	0.815
160	R-Pallidum	0.866	99-147	R-G_pariet_inf-Angular - R-S_temporal_sup	0.453	73	Putamen_L	0.581	31-71	Cingulum_Ant_L - Caudate_L	0.801
154	L-Pallidum	0.820	38-73	L-G_temporal_middle - L-S_temporal_sup	0.412	20	Supp_Motor_Area_R	0.492	20-58	Supp_Motor_Area_R - Postcentral_R	0.698
159	R-Putamen	0.814	112-147	R-G_temporal_middle - R-S_temporal_sup	0.411	78	Thalamus_R	0.489	77-99	Thalamus_L - Cerebellum_6_L	0.662
152	L-Thalamus_Proper	0.745	29-45	L-G_precentral - L-S_central	0.393	11	Frontal_Inf_Oper_L	0.479	32-72	Cingulum_Ant_R - Caudate_R	0.634
151	L-Hippocampus	0.556	28-45	L-G_postcentral - L-S_central	0.379	76	Pallidum_R	0.444	19-57	Supp_Motor_Area_L - Postcentral_L	0.576
149	L-Amygdala	0.554	15-52	L-G_front_middle - L-S_front_inf	0.377	91	Cerebellum_Crus1_L	0.438	4-38	Frontal_Sup_R - Hippocampus_R	0.484
155	R-Amygdala	0.551	89-127	R-G_front_middle - R-S_front_middle	0.362	75	Pallidum_L	0.421	33-78	Cingulum_Mid_L - Thalamus_R	0.411
157	R-Hippocampus	0.550	103-119	R-G_precentral - R-S_central	0.347	70	Paracentral_Lobule_R	0.418	30-82	Insula_R - Temporal_Sup_R	0.400

are given in Sec. E) We evaluated all models using 5-fold cross-validation with 10 different random seeds and report accuracy, macro-averaged precision, and macro-averaged recall.

5.2 Quantitative Results

Comparison with baselines. As shown in Tab. 1, Ours (SSWT w/ ScaleX) achieves the best mean performance across all datasets and metrics, including recall (i.e., sensitivity), which is particularly important for disease diagnosis. Compared with AGT, the strongest baseline, our model improves accuracy, precision, and recall by 3.3%p, 3.3%p, and 2.5%p on average across the three datasets. To account for the dependence among repeated cross-validation runs, we assess the accuracy improvement over AGT using the corrected resampled t-test (Nadeau & Bengio, 1999; Bouckaert & Frank, 2004) (5 folds $\times$ 10 random seeds). The improvement is statistically significant on ADNI-CT ($p = 0.004$), ADNI-FDG ($p = 0.021$), and PPMI ($p = 0.013$), with all three comparisons remaining significant after Holm–Bonferroni correction. These results demonstrate consistent improvements over existing methods across the evaluated datasets.

Table 3: Zero-shot evaluation (PPMI $\rightarrow$ TaoWu, Neurocon)

Model	PPMI $\rightarrow$ TaoWu			PPMI $\rightarrow$ Neurocon		
	Acc. (%)	Pre. (%)	Rec. (%)	Acc. (%)	Pre. (%)	Rec. (%)
SaNN	60.6 ± 8.9	62.7 ± 4.1	60.6 ± 8.9	65.2 ± 3.4	64.9 ± 2.3	66.0 ± 4.3
ALTER	54.5 ± 1.9	65.1 ± 6.0	54.5 ± 1.9	63.8 ± 3.2	58.8 ± 7.6	55.3 ± 4.4
ContrastPool	64.4 ± 0.7	68.5 ± 7.2	64.4 ± 0.7	62.3 ± 2.7	64.9 ± 5.2	65.6 ± 9.2
AGT	70.6 ± 6.3	71.6 ± 5.9	70.6 ± 6.3	73.2 ± 4.2	72.8 ± 3.9	74.5 ± 5.7
Ours w/o ScaleX	73.1 ± 2.9	76.2 ± 4.4	73.1 ± 2.9	74.7 ± 1.3	75.0 ± 6.3	76.9 ± 3.9
Ours	$\mathbf{81.8 \pm 2.6}$	$\mathbf{83.9 \pm 3.0}$	$\mathbf{81.8 \pm 2.6}$	$\mathbf{82.3 \pm 3.6}$	$\mathbf{81.2 \pm 2.4}$	$\mathbf{83.2 \pm 3.1}$

Zero-shot results. To evaluate robustness under distribution shift, we perform zero-shot evaluation on external Parkinson's disease datasets (TaoWu and Neurocon cohorts from (Xu et al., 2023)). Each model is trained on the PPMI dataset and directly evaluated on these datasets without fine-tuning. These datasets contain fewer subjects than PPMI and are independently collected, and thus serve as out-of-distribution test sets. TaoWu ($N = 40$) and Neurocon ($N = 41$) are used exclusively as external test cohorts and differ from PPMI in both acquisition site and PD definition; thus, the evaluation reflects robustness under combined cross-cohort distribution shifts. As shown in Tab. 3, baselines and ours w/o ScaleX suffer substantial performance degradation under this setting. In contrast, our full model (SSWT w/ ScaleX) consistently maintains strong performance across both datasets, indicating improved robustness under distribution shift. These findings further support the effectiveness of adaptive scale selection for cross-cohort generalization.

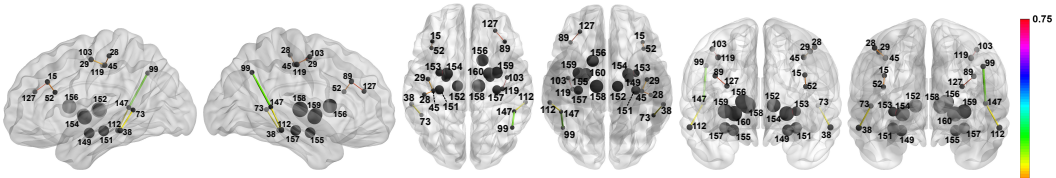
Table 4: Ablation study on the SSWT for the ADNI-CT.

Use of signals		Use of filters		Acc. (%)	Pre. (%)	Rec. (%)
Node	Edge	Low pass	Band pass			
✓	×	✓	✓	91.8 ± 0.9	91.4 ± 1.0	93.1 ± 1.1
×	✓	✓	✓	90.9 ± 1.5	91.0 ± 1.9	91.5 ± 1.5
✓	✓	✓	✓	$\mathbf{94.9 \pm 1.3}$	$\mathbf{95.4 \pm 0.8}$	$\mathbf{94.4 \pm 1.7}$
✓	✓	×	✓	94.4 ± 1.3	94.0 ± 1.4	95.2 ± 1.1
✓	✓	✓	×	92.3 ± 1.1	91.7 ± 1.4	93.3 ± 0.9

Ablation Study. To assess the effect of node and edge signals and different filter types in SSWT, we conduct an ablation study on the ADNI-CT (Tab. 4). Using both node and edge signals yields the highest accuracy, confirming their complementary roles in capturing structural properties of brain networks. Removing either signal degrades performance, emphasizing their contributions to regional characterization and connectivity. For filter selection, combining low-pass and band-pass filters achieves the best result, balancing global and localized information. Excluding either filter reduces performance, highlighting their roles in capturing fine-grained spectral features and global structure. Further analyses of expert configurations and routing balance are in Sec. F.

5.3 Model Behavior Analysis

Interpretation. To provide clinical interpretability, we investigate the class-averaged Grad-CAM (Selvaraju et al., 2017) of nodes and edges in classifying AD and PD. Tab. 2 reports the top-10 ROIs and connectomes with the highest gradient activation for ADNI-CT and PPMI experiments.



(a) (b) (c) (d) (e) (f)

Figure 3: Top-10 activated ROIs and connectomes for ADNI-CT classification. (a)/(b): outer view of left/right hemisphere, (c)/(d): top/bottom, (e)/(f): front/rear view. Node size and edge color/thickness indicate activations, with node indices corresponding to the Destrieux atlas.

For ADNI-CT, the results are visualized in Fig. 3. All highly activated nodes are located within subcortical structures (e.g., hippocampus, amygdala), which are crucial for memory, cognitive processing, and motor functions, and are strongly associated with AD (Tentolouris-Piperas et al., 2017; Shukla et al., 2024). The thalamus exhibits the highest activation, aligning with evidence that its atrophy correlates with cognitive decline (de Jong et al., 2008). The selected edges connect both subcortical and cortical regions, particularly in the temporal and frontal lobes, which play essential roles in memory, language, and executive functions (Palmer et al., 1987; Killiany et al., 1993).

For PPMI, the putamen, pallidum, and caudate, key structures of the basal ganglia, exhibited high activation in nodes and edges across both hemispheres. These regions, essential for motor control, cognition, and emotion, are strongly linked to PD (de la Fuente-Fernández, 2013; Obeso et al., 2000). Additionally, strong activation is observed in the putamen-thalamus and thalamus-cerebellum connections, consistent with known PD-related disruptions in the basal ganglia-thalamic and cerebello-thalamo-cortical circuits (Helmich et al., 2012). Across both datasets, our model consistently captures bilaterally symmetric activation patterns, reinforcing its classification reliability.

Class-Specific Expert Selection. The ScaleX dynamically selects spectral experts based on the structural properties of each brain network. To analyze class-wise expert selection, we visualize the frequency of node expert activations across diagnostic groups in ADNI-CT, using five experts and $k = 2$ (Fig. 4). Although the routing balance regularization encourages balanced expert usage overall, selection patterns vary significantly across classes. Within each class, the model exhibits distinct activation preference, highlighting that the routing mechanism effectively differentiates disease stages. These class-specific patterns indicate that ScaleX dynamically adjusts its scale selection to capture the most relevant spectral features for each neurodegeneration stage.



(a) CN (b) SMC (c) EMCI (d) LMCI (e) AD

Figure 4: Expert selection patterns across different diagnostic groups in the ADNI-CT dataset. Each heatmap illustrates expert activation frequencies for node-level representations, with 5 experts and a top-k value of 2.

6 Conclusion

In this paper, we proposed a novel framework that extends the spectral graph wavelet to simplicial complexes, enabling multi-scale spectral analysis of both node and edge features in a graph, while adaptively capturing heterogeneous spectral characteristics across nodes and edges. Leveraging the Hodge Laplacian, SSWT adaptively filters graph signals on both nodes and edges, and ScaleX extends it to multi-resolution learning, enhancing the spectral representation of complex structures. Extensive experiments on multiple clinical benchmark datasets, including ADNI and PPMI, show that our model achieves state-of-the-art prediction performance with clinically meaningful interpretation, demonstrating its potential for brain network analysis in neurodegeneration as well as broader applications in both neuroscience and graph analysis.

AI USE STATEMENT

We used Large Language Models (LLMs) solely to aid in polishing the writing of this manuscript. No LLMs were used for generating research ideas, experimental design, data analysis, or any other scientific contributions.

REPRODUCIBILITY

To ensure the reproducibility of our work, we provide comprehensive implementation details, including model architectures, optimization settings, and baseline tuning protocols, in Sec. E. The datasets used in our experiments, ADNI and PPMI, are publicly available, and the full preprocessing pipeline is described in Sec. D. For the theoretical component, all mathematical assumptions and proofs are explicitly provided in Sec. A. All experiments are conducted with 5-fold cross-validation, and results are reported as mean values with standard deviations. We will also release the full code and setup to facilitate reproducibility of our work.

ETHICS

This study uses only publicly available neuroimaging datasets, ADNI and PPMI, which are distributed for research use. No additional data involving human subjects were collected or used. The datasets are widely adopted benchmarks in the neuroimaging community and contain no personally identifiable information, minimizing privacy concerns.

REFERENCES

D Vijay Anand and Moo K Chung. Hodge Laplacian of brain networks. *IEEE transactions on medical imaging*, 42(5):1563–1573, 2023.

Remco R Bouckaert and Eibe Frank. Evaluating the replicability of significance tests for comparing learning algorithms. In *Pacific-Asia conference on knowledge discovery and data mining*, pp. 3–12. Springer, 2004.

Edward T Bullmore et al. Brain graphs: graphical models of the human brain connectome. *Annual review of clinical psychology*, 7(1):113–140, 2011.

Lei Cai, Jundong Li, Jie Wang, and Shuiwang Ji. Line graph neural networks for link prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5103–5113, 2021.

Ke Chen, Lei Xu, et al. Improved learning algorithms for mixture of experts in multiclass classification. *Neural networks*, 12(9):1229–1252, 1999.

Hyuna Cho, Jaeyoon Sim, et al. Neurodegenerative brain network classification via adaptive diffusion with temporal regularization. In *International Conference on Machine Learning (ICML)*, 2024.

Injun Choi, Guorong Wu, et al. How much to aggregate: Learning adaptive node-wise scales on graphs for brain networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 376–385. Springer, 2022.

O Ciccarelli, TE Behrens, et al. Probabilistic diffusion tractography: a potential tool to assess the rate of disease progression in amyotrophic lateral sclerosis. *Brain*, 129(7):1859–1871, 2006.

Sixtus Dakurah, D Vijay Anand, et al. Modelling cycles in brain networks with the Hodge Laplacian. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 326–335. Springer, 2022.

Laura W de Jong, Karin van der Hiele, et al. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain*, 131(12):3277–3285, 2008.

Raúl de la Fuente-Fernández. Imaging of dopamine in PD and implications for motor and neuropsychiatric manifestations of PD. *Frontiers in neurology*, 4:90, 2013.

Michaël Defferrard, Xavier Bresson, et al. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in Neural Information Processing Systems (NeurIPS)*, 29, 2016.

Christophe Destrieux, Bruce Fischl, et al. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1–15, 2010.

William Fedus, Barret Zoph, et al. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022.

Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

Johannes Gasteiger, Stefan Weißenberger, et al. Diffusion improves graph learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.

Zihao Guo, Qingyun Sun, et al. Graphmore: Mitigating topological heterogeneity via mixture of riemannian experts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 11754–11762, 2025.

Sravanthi Gurugubelli and Sundeep Prabhakar Chepuri. Sann: Simple yet powerful simplicial-aware neural networks. In *International Conference on Learning Representations (ICLR)*, 2024.

David K Hammond, Pierre Vandergheynst, et al. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.

Ran He, Karthik Gopinath, et al. Spectral graph transformer networks for brain surface parcellation. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 372–376. IEEE, 2020.

Anibal Sólon Heinsfeld, Alexandre Rosa Franco, et al. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: clinical*, 17:16–23, 2018.

Rick C Helmich, Mark Hallett, et al. Cerebral causes and consequences of parkinsonian resting tremor: a tale of two circuits? *Brain*, 135(11):3206–3226, 2012.

Weihua Hu, Matthias Fey, et al. Open graph benchmark: Datasets for machine learning on graphs. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:22118–22133, 2020.

Robert A Jacobs, Michael I Jordan, et al. Adaptive mixtures of local experts. *Neural computation*, 3(1):79–87, 1991.

Michael Jordan, Zoubin Ghahramani, et al. Hidden markov decision trees. *Advances in Neural Information Processing Systems (NeurIPS)*, 9, 1996.

Michael I Jordan et al. Hierarchical mixtures of experts and the em algorithm. *Neural computation*, 6(2):181–214, 1994.

Xuan Kan, Wei Dai, et al. Brain network transformer. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:25586–25599, 2022.

Ronald J Killiany, Mark B Moss, et al. Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. *Archives of neurology*, 50(9):949–954, 1993.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations (ICLR)*, 2017.

Xiaoxiao Li, Yuan Zhou, et al. BrainGNN: Interpretable brain graph neural network for fMRI analysis. *Medical image analysis*, 74:102233, 2021.

Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

Kenneth Marek, Danna Jennings, et al. The Parkinson progression marker initiative (PPMI). *Progress in neurobiology*, 95(4):629–635, 2011.

Christopher Morris, Nils M Kriege, et al. TUDataset: A collection of benchmark datasets for learning with graphs. *arXiv preprint arXiv:2007.08663*, 2020.

Susanne G Mueller, Michael W Weiner, et al. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 15(4):869, 2005.

Basil Mustafa, Carlos Riquelme, et al. Multimodal contrastive learning with limoe: the language-image mixture of experts. *Advances in Neural Information Processing Systems (NeurIPS)*, 35: 9564–9576, 2022.

Claude Nadeau and Yoshua Bengio. Inference for the generalization error. *Advances in neural information processing systems*, 12, 1999.

José A Obeso, Maria C Rodriguez-Oroz, et al. Pathophysiology of the basal ganglia in Parkinson's disease. *Trends in neurosciences*, 23:S8–S19, 2000.

Alan M Palmer, Gordon K Wilcock, et al. Monoaminergic innervation of the frontal and temporal lobes in Alzheimer's disease. *Brain research*, 401(2):231–238, 1987.

Michela Pievani, Nicola Filippini, et al. Brain connectivity in neurodegenerative diseases—from phenotype to proteinopathy. *Nature Reviews Neurology*, 10(11):620–633, 2014.

Ashish Raj, Amy Kuceyeski, et al. A network diffusion model of disease progression in dementia. *Neuron*, 73(6):1204–1215, 2012.

Carlos Riquelme, Joan Puigcerver, et al. Scaling vision with sparse mixture of experts. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:8583–8595, 2021.

T Mitchell Roddenberry, Florian Frantzen, et al. Hodgelets: Localized spectral representations of flows on simplicial complexes. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5922–5926. IEEE, 2022.

Stephen Roller, Sainbayar Sukhbaatar, et al. Hash layers for large sparse models. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:17555–17566, 2021.

Michael T Schaub, Austin R Benson, et al. Random walks on simplicial complexes and the normalized Hodge 1-Laplacian. *SIAM Review*, 62(2):353–391, 2020.

Ramprasaath R Selvaraju, Michael Cogswell, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *International Conference on Computer Vision (ICCV)*, pp. 618–626, 2017.

Noam Shazeer, Mirhoseini, et al. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *International Conference on Learning Representations (ICLR)*, 2017.

Yuge Shi, Brooks Paige, et al. Variational mixture-of-experts autoencoders for multi-modal deep generative models. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.

Amar Shukla, Rajeev Tiwari, and Shamik Tiwari. Analyzing subcortical structures in Alzheimer's disease using ensemble learning. *Biomedical Signal Processing and Control*, 87:105407, 2024.

David I Shuman, Sunil K Narang, et al. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.

Jaeyoon Sim, Sooyeon Jeon, et al. Learning to approximate adaptive kernel convolution on graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4882–4890, 2024.

Vasileios Tentolouris-Piperas, Natalie S Ryan, et al. Brain imaging evidence of early involvement of subcortical regions in familial and sporadic Alzheimer's disease. *Brain research*, 1655:23–32, 2017.

Betty M Tijms, Alle Meije Wink, et al. Alzheimer's disease: connecting findings from graph theoretical studies of brain networks. *Neurobiology of aging*, 34(8):2023–2036, 2013.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

Petar Veličković, Guillem Cucurull, et al. Graph attention networks. In *International Conference on Learning Representations (ICLR)*, 2018.

Prashanthi Vemuri and Clifford R Jack. Role of structural MRI in Alzheimer's disease. *Alzheimer's research & therapy*, 2:1–10, 2010.

Haotao Wang, Ziyu Jiang, et al. Graph mixture of experts: Learning on large-scale graphs with explicit diversity modeling. *Advances in Neural Information Processing Systems (NeurIPS)*, 36: 50825–50837, 2023.

Bingbing Xu, Huawei Shen, et al. Graph convolutional networks using heat kernel for semi-supervised learning. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2019.

Jiaxing Xu, Yunhan Yang, et al. Data-driven network neuroscience: On data collection and benchmark. *Advances in Neural Information Processing Systems (NeurIPS)*, 36:21841–21856, 2023.

Jiaxing Xu, Qingtian Bian, et al. Contrastive graph pooling for explainable classification of brain networks. *IEEE Transactions on Medical Imaging*, 2024.

Maosheng Yang, Elvin Isufi, et al. Simplicial convolutional neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8847–8851. IEEE, 2022.

Shuo Yu, Shan Jin, et al. Long-range brain graph transformer. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:24472–24495, 2024.

Eduard Zamfir, Zongwei Wu, Nancy Mehta, et al. See more details: Efficient image super-resolution by experts mining. In *International Conference on Machine Learning (ICML)*, 2024.

Jialin Zhao, Yuxiao Dong, et al. Adaptive diffusion in graph neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:23321–23333, 2021.

APPENDIX

In the appendix, we provide supplementary materials to support the main manuscript. Specifically, we include: **1)** proofs of Propositions regarding the unified spectral representation for Hodge Laplacians and the orientation equivariance of our edge filtering; **2)** discussion on extending SSWT to higher-order simplicial complexes; **3)** a rationale for wavelet kernel selection; **4)** detailed description of the ADNI and PPMI datasets used in our experiments; **5)** full implementation details and hyperparameter settings for our model, along with a comprehensive baseline tuning protocols; **6)** additional experimental results on the ABIDE dataset for evaluating robustness beyond neurodegenerative disease settings; **7)** an ablation study on the number of experts and top-k selection in ScaleX; **8)** the effectiveness of routing balance regularization on model performance; **9)** a discussion on model complexity; and **10)** a discussion of the limitations and broader societal impacts of the proposed method.

A PROOF OF PROPOSITIONS

Proposition 1. *Let $B_1 = U\Sigma V^T$ be the singular value decomposition (SVD) of the incidence matrix. Then, the eigenvectors of Hodge Laplacians $\mathcal{L}_0$ and $\mathcal{L}_1$ are given by U and V, respectively (i.e., $U_0 = U, U_1 = V$), and excluding the zero eigenvalues, both Laplacians share the same non-zero eigenvalues $\lambda_i = \sigma_i^2$, where σ_i are the non-zero singular values from Σ.*

Proof. Substituting the SVD $B_1 = U\Sigma V^T$ into the definitions $\mathcal{L}_0 = B_1 B_1^T$ and $\mathcal{L}_1 = B_1^T B_1$, and utilizing the orthogonality properties $U^T U = I$ and $V^T V = I$, we derive:

$$\begin{aligned} \mathcal{L}_0 &= (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma \underbrace{V^T V}_{=I} \Sigma^T U^T = U(\Sigma\Sigma^T)U^T, \\ \mathcal{L}_1 &= (U\Sigma V^T)^T(U\Sigma V^T) = V\Sigma^T \underbrace{U^T U}_{=I} \Sigma V^T = V(\Sigma^T\Sigma)V^T. \end{aligned} \tag{1}$$

Since $\Sigma\Sigma^T$ and $\Sigma^T\Sigma$ are diagonal matrices sharing the same non-zero entries σ_i^2, these derivations confirm that $\mathcal{L}_0$ and $\mathcal{L}_1$ share the non-zero eigenvalues, with U and V serving as their respective eigenvector bases. □

Proposition 2. *Consider a boundary matrix $B_1 \in \mathbb{R}^{N\times M}$ with arbitrary orientations and edge signal ω of a graph $\mathcal{G}$. Let $J \in \mathbb{R}^{M\times M}$ be an arbitrary diagonal matrix encoding changes in edge orientation, where $J[i,i] = -1$ if the orientation of the i-th edge is flipped, and $J[i,i] = 1$ otherwise. Define the reoriented graph $\mathcal{G}'$, with $\omega' = J\omega$ and $B_1' = B_1 J$. Then, the filtered edge signal ω_t' on $\mathcal{G}'$ satisfies $\omega_t' = J\omega_t$, demonstrating that the ISWT is equivariant to edge orientation.*

Proof. Let $J \in \mathbb{R}^{M\times M}$ be a diagonal matrix defined as:

$$J[i,i] = \begin{cases} -1, & \text{if the } i\text{-th edge is flipped,} \\ 1, & \text{otherwise.} \end{cases} \tag{2}$$

Altering the orientation of edges corresponds to flipping the signs of the corresponding columns in B_1 as $B_1' = B_1 J$. The modified Hodge Laplacian $\mathcal{L}_1'$ follows as:

$$\mathcal{L}_1' = B_1'^T B_1' = (B_1 J)^T (B_1 J) = J^T B_1^T B_1 J = J^T \mathcal{L}_1 J. \tag{3}$$

Since J is diagonal with entries ± 1, it satisfies $J = J^T = J^{-1}$. Thus, the relationship between $\mathcal{L}_1'$ and $\mathcal{L}_1$ represents a similarity transformation. This implies that they share the same eigenvalues Λ_1. Let the eigendecomposition of the original Laplacian be $\mathcal{L}_1 = U_1\Lambda_1 U_1^T$. The eigenvectors of the reoriented graph, denoted by U_1', are related to U_1 as follows:

$$\mathcal{L}_1'(JU_1) = (J\mathcal{L}_1 J^{-1})JU_1 = J\mathcal{L}_1 U_1 = JU_1\Lambda_1. \tag{4}$$

From Eq. (4), we identify that $U_1' = JU_1$. Since $J^{-1} = J$, it also holds that $U_1 = JU_1'$.

Now, consider the wavelet coefficients $W_\omega(t) = \phi_t \cdot \omega = U_1 \mathcal{K}(t\Lambda_1) U_1^T \omega$. Applying the transformation $U_1' = JU_1$ and $\omega' = J\omega$, the coefficients for the reoriented graph are:

$$W'_\omega(t) = U_1' \mathcal{K}(t\Lambda_1) U_1'^T \omega' \tag{5}$$

$$= (JU_1)\mathcal{K}(t\Lambda_1)(JU_1)^T (J\omega) \tag{6}$$

$$= JU_1 \mathcal{K}(t\Lambda_1) U_1^T J^T J\omega \tag{7}$$

$$= JU_1 \mathcal{K}(t\Lambda_1) U_1^T \omega = JW_\omega(t). \tag{8}$$

Similarly, the filtered edge signal $\omega_t = U_1 \mathcal{K}^2(t\Lambda_1) U_1^T \omega$ transforms as:

$$\omega_t' = U_1' \mathcal{K}^2(t\Lambda_1) U_1'^T \omega' \tag{9}$$

$$= (JU_1)\mathcal{K}^2(t\Lambda_1)(JU_1)^T (J\omega) \tag{10}$$

$$= JU_1 \mathcal{K}^2(t\Lambda_1) U_1^T J^T J\omega \tag{11}$$

$$= JU_1 \mathcal{K}^2(t\Lambda_1) U_1^T \omega = J\omega_t. \tag{12}$$

This demonstrates that flipping the orientation of edges only changes the sign of the corresponding components of $W_\omega(t)$ and ω_t, while preserving their magnitudes. This proves orientation equivariance. □

B Extension to Higher-Order Simplicial Complexes

The current implementation of SSWT treats a brain connectome as a 1-dimensional simplicial complex, where nodes and edges correspond to 0- and 1-simplices, respectively. In this setting, no higher-order simplices are explicitly included. However, when higher-order relations are available or constructed from data, the graph can be lifted to a higher-order simplicial complex. For example, one may build a clique complex by filling triangles formed by 3-cliques, or incorporate domain-defined group interactions as 2- or higher-dimensional simplices. Our main experiments do not use such higher-order constructions, but the proposed spectral filtering framework is compatible with this general setting.

For a 1-dimensional simplicial complex, the edge-level Hodge Laplacian reduces to

$$\mathcal{L}_1 = B_1^T B_1, \tag{13}$$

since no higher-order simplices exist and thus $B_2 = 0$. When higher-order simplices are introduced, the corresponding upper-adjacency term becomes nonzero, yielding

$$\mathcal{L}_1 = B_1^T B_1 + B_2 B_2^T. \tag{14}$$

Here, the lower term $B_1^T B_1$ captures relationships between edges sharing nodes, while the upper term $B_2 B_2^T$ captures higher-order edge interactions induced by shared triangles. More generally, the r-th Hodge Laplacian is defined as

$$\mathcal{L}_r = B_r^T B_r + B_{r+1} B_{r+1}^T, \tag{15}$$

which encodes interactions between neighboring r-simplices through shared lower-dimensional faces and higher-dimensional cofaces. Since the combinatorial Hodge Laplacian $\mathcal{L}_r$ is symmetric positive semi-definite, it admits the eigendecomposition

$$\mathcal{L}_r = U_r \Lambda_r U_r^T. \tag{16}$$

Accordingly, the same spectral filtering construction used in the main text can be applied to arbitrary simplex orders as

$$\psi_s^{(r)} = U_r \mathcal{K}(s\Lambda_r) U_r^T, \tag{17}$$

where $\mathcal{K}(\cdot)$ is a spectral wavelet kernel and s is a scale parameter. For an r-simplex signal $x^{(r)}$, the corresponding scale-filtered signal is given by

$$x_s^{(r)} = U_r \mathcal{K}^2(s\Lambda_r) U_r^T x^{(r)}. \tag{18}$$

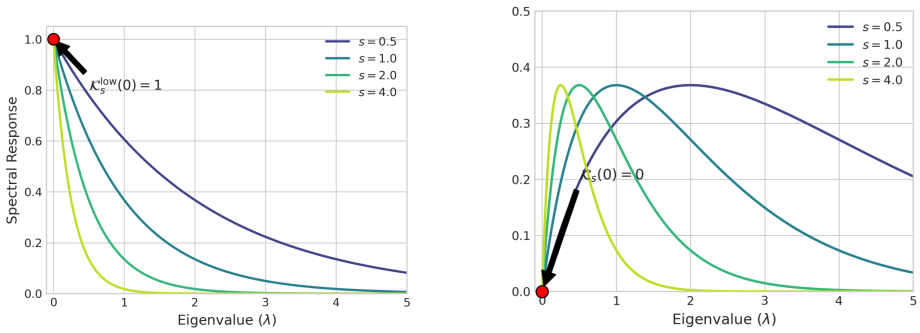
Unlike graph spectral operators that only encode pairwise connectivity, higher-order Hodge Laplacians additionally capture interactions induced by higher-dimensional simplices. Consequently, the proposed spectral filtering framework can represent multi-scale interactions beyond pairwise graph relations whenever higher-order simplicial structures are available.

C Rationale for Kernel Selection

We employ parametric kernels that strictly follow the generating principles of Spectral Graph Wavelet Theory (Hammond et al., 2011). According to the theory, a wavelet kernel $g(\lambda)$ must satisfy the admissibility condition $g(0) = 0$, which implies that it inherently suppresses the low-frequency signal components at $\lambda = 0$ (i.e., the global average or DC component). To ensure a complete spectral representation, it is necessary to introduce a scaling function $h(\lambda)$ that satisfies $h(0) > 0$ to capture these missing low-frequency features. Our selected kernels are designed to satisfy these complementary roles explicitly:

Band-pass Kernel (Wavelet): We use $\mathcal{K}_s(\lambda) = s\lambda e^{-s\lambda}$ (Fig. 1b). Since this function satisfies $\mathcal{K}_s(0) = 0$ and decays as $\lambda \to \infty$, it functions as a valid band-pass filter, effectively capturing structural transitions and intermediate frequencies while ignoring the global mean.

Low-pass Kernel (Scaling Function): We adopt the heat kernel $\mathcal{K}_s^{\text{low}}(\lambda) = e^{-s\lambda}$ (Fig. 1a). Unlike the wavelet kernel, this function satisfies $\mathcal{K}_s^{\text{low}}(0) = 1$, allowing it to act as a scaling function that explicitly captures the global structural information and low-frequency trends that the band-pass kernels miss.



(a) Low-pass kernel: $\mathcal{K}_s^{\text{low}}(\lambda) = e^{-s\lambda}$ (b) Band-pass kernel: $\mathcal{K}_s(\lambda) = s\lambda e^{-s\lambda}$

Figure 1: Frequency response of the chosen spectral kernels. **(a)** The low-pass kernel $\mathcal{K}_s^{\text{low}}(\lambda)$ satisfies $\mathcal{K}_s^{\text{low}}(0) = 1$ (red dot), allowing it to capture global structural information and the DC component (global mean) that band-pass filters miss. **(b)** The band-pass kernels $\mathcal{K}_s(\lambda)$ satisfy the admissibility condition $\mathcal{K}_s(0) = 0$ (red dot). They vanish at the origin and effectively isolate intermediate frequency features at different scales s. Together, they provide a complete coverage of the graph spectrum.

D Detailed Dataset Description

We utilize two neurodegenerative brain network datasets: the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Parkinson's Progression Markers Initiative (PPMI). For both datasets, each observation corresponds to a unique subject. For participants with multiple longitudinal visits, only the earliest available visit is retained, resulting in one observation per subject.

D.1 ADNI

The ADNI study (Mueller et al., 2005) is a publicly available dataset designed for Alzheimer's Disease (AD) research, providing multimodal imaging and biomarker data across various stages of cognitive decline. Following prior studies (Choi et al., 2022; Sim et al., 2024; Cho et al., 2024), the structural brain graphs in the ADNI dataset are constructed through a multi-step imaging pipeline. First, skull stripping and tissue segmentation are performed on T1-weighted MRI scans. The brain is then parcellated into 160 anatomical regions of interest (ROIs) based on the Destrieux atlas (Destrieux et al., 2010). Diffusion tensors are estimated from DWI data, and cortical surfaces are reconstructed using FreeSurfer (Fischl, 2012). Probabilistic tractography is applied to extract white matter fiber tracts between ROIs, from which the graph connectivity is defined and tract counts are used as edge signals, reflecting the integrity of structural connectivity (Ciccarelli et al., 2006). The incidence matrix B_1 is constructed from the resulting connectivity, where each retained edge is assigned an arbitrary orientation and its corresponding column contains one -1 and one $+1$ at the incident ROIs. Finally, region-wise imaging features are extracted: cortical thickness (CT) from MRI and standardized uptake value ratios (SUVRs) from FDG-PET, where the cerebellum is used as the reference region for SUVR normalization. These modalities are widely used in Alzheimer's Disease (AD) research, with cortical thickness capturing gray matter atrophy and FDG-PET reflecting metabolic deficits (Vemuri & Jack, 2010).

The dataset consists of five diagnostic groups: Cognitively Normal (CN), Significant Memory Concern (SMC), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD). The demographic distribution of the ADNI dataset is summarized in Tab. 1. It includes a total of 1,323 subjects, with 359 CN, 181 SMC, 437 EMCI, 180 LMCI, and 166 AD cases. The average age within these groups ranges from 70.9 to 74.8 years, with standard deviations between 1.4 and 8.7 years. Gender distribution varies across groups, with CN comprising 178 males and 181 females, while AD includes 102 males and 64 females.

Table 1: Demographics of the ADNI dataset.

Biomarker	Category	CN	SMC	EMCI	LMCI	AD
Cortical Thickness & FDG	# of subjects	359	181	437	180	166
	Gender (M / F)	178 / 181	69 / 112	249 / 188	119 / 61	102 / 64
	Age (Mean±Std)	72.8±1.4	72.0±5.2	71.0±7.9	70.9±6.1	74.8±8.7

D.2 PPMI

The PPMI dataset (Marek et al., 2011) focuses on understanding the progression of Parkinson's Disease (PD) through imaging and biomarker analysis. Functional brain networks are constructed by parcellating resting-state fMRI into 116 regions based on the AAL atlas (Tzourio-Mazoyer et al., 2002). Node features are represented by ROI-wise Blood-Oxygen-Level-Dependent (BOLD) time series, with longer sequences truncated to the shortest sequence length in the dataset. Edge features represent functional connectivity and are computed as pairwise Pearson correlations between the ROI-wise BOLD signals (Xu et al., 2023). Only positive correlations greater than 0.5 are retained, while negative and weaker correlations are set to zero. Using a fixed indexing over candidate ROI pairs, retained edges contain their correlation values, whereas absent edges are represented by zero. The incidence matrix B_1 follows the same indexing, where each retained edge is assigned an arbitrary orientation with one -1 and one $+1$ at its incident ROIs, while columns corresponding to absent edges are set to zero.

The dataset consists of three diagnostic groups: Cognitively Normal (CN), Prodromal, and Parkinson's Disease (PD), representing different stages of disease progression. The SWEDD group (Scans

Without Evidence for Dopaminergic Deficit) was excluded from our study. Although SWEDD patients present parkinsonian motor symptoms, they show normal presynaptic dopaminergic imaging findings, making them pathophysiologically distinct from PD. Including them could introduce confounding effects by acting as outliers and hindering the model's ability to capture genuine disease progression signals. The official PPMI Data User Guide[1] also describes SWEDD as a small legacy cohort that researchers may wish to exclude depending on their research purpose, which aligns with our decision. The demographic details of the PPMI dataset are presented in Tab. 2. It includes a total of 195 subjects, with 15 CN, 67 Prodromal, and 113 PD cases. The average age within these groups ranges from 62.0 to 64.3 years, with standard deviations up to 10.0 years. Gender distribution also varies, with CN consisting of 12 males and 3 females, while the PD group comprises 77 males and 36 females.

Table 2: Demographics of the PPMI dataset.

Biomarker	**Category**	**CN**	**Prodromal**	**PD**
BOLD	# of subjects	15	67	113
	Gender (M / F)	12 / 3	38 / 29	77 / 36
	Age (Mean±Std)	62.0±10.0	64.3±8.7	64.0±9.5

[1] https://www.ppmi-info.org/sites/default/files/docs/PPMI%20Data%20User%20Guide.pdf

E IMPLEMENTATION DETAILS

All models, including our model and baseline methods, were implemented using PyTorch and trained on a single NVIDIA RTX 6000 Ada Generation GPU. We used the Adam optimizer and trained each model for up to 5000 epochs with early stopping. Node and edge signals were normalized independently within each subject graph by centering and range scaling using only the values from that graph. The Hodge Laplacians and their spectral decompositions were also computed independently for each subject graph.

To ensure robust evaluation, all experiments were repeated using label-stratified 5-fold cross-validation over 10 random seeds. In each fold, four folds (80%) were used as the development set, while the remaining fold (20%) was held out as the test set. The development set was further divided into training and validation subsets using a label-stratified 4:1 split, resulting in approximate train/validation/test proportions of 64%/16%/20%. Early stopping and checkpoint selection were based exclusively on the validation loss with a patience of 50 epochs, and the checkpoint with the lowest validation loss was retained.

A grid search was conducted over the following common hyperparameter ranges: the number of hidden units in $\{2, 4, 6, 8, 16, 32, 64\}$, dropout rates in $\{0.2, 0.4, 0.6, 0.8\}$, and learning rates in $\{0.1, 0.01, 0.001, 0.0001\}$. Hyperparameters were selected based on validation performance, without using the test-set performance for model selection. No class weighting was applied to either our model or the baseline methods.

Our Settings. Tab. 3 summarizes the hyperparameters used for our model across datasets. The batch size was set to include all samples in the training subset of each fold. Dropout rates were set to 0.6 for ADNI-CT and ADNI-FDG, and 0.4 for PPMI. The learning rates were set to 1×10^{-3} for ADNI-CT and ADNI-FDG, and 2×10^{-2} for PPMI. The weight for λ_{bal} which controls the contribution of the load balancing term in the overall objective was set to 5 across all datasets. The hidden dimension h was set to 6 for ADNI-CT, 4 for ADNI-FDG, and 16 for PPMI, and we used 2 readout layers across all datasets. For graph-level prediction, the combined node–edge embedding $[\tilde{\mathbf{f}}; \boldsymbol{\omega}^{\text{align}}]$ is flattened using the fixed atlas-based ROI ordering and the corresponding fixed edge indexing, and then passed to the readout function $F(\cdot)$. The number of node experts P and edge experts Q for ScaleX was set to (5, 5) for ADNI-CT, (5, 4) for ADNI-FDG, and (5, 5) for PPMI. We used $k = 2$ for sparse expert selection.

For external evaluation on TaoWu and Neurocon, predictions are obtained by comparing the CN and PD outputs of the three-class classifier trained on PPMI, while excluding the Prodromal output. No fine-tuning is performed on the target cohorts.

Table 3: Hyperparameters of our model for experiments.

Hyperparameter	**ADNI-CT**	**ADNI-FDG**	**PPMI**
Optimizer	Adam	Adam	Adam
Number of epochs	5000	5000	5000
Weight decay	5×10^{-4}	5×10^{-4}	5×10^{-4}
Dropout rate	0.6	0.6	0.4
Learning rate	1×10^{-3}	1×10^{-3}	2×10^{-2}
Weight for λ_{bal}	5	5	5
Hidden dimension h	6	4	16
Number of $F(\cdot)$ layers	2	2	2
Number of experts (P, Q)	(5, 5)	(5, 4)	(5, 5)
k for top-k selection	2	2	2

Baseline-specific Settings. To ensure fair and reliable comparisons, we used official open-source implementation for each method. In addition to the shared grid search described above, we further tuned model-specific key hyperparameters to adapt each method to our domain-specific datasets:

- **GDC** (Gasteiger et al., 2019): We evaluated both the Heat Kernel and Personalized PageRank (PPR) variants. For each, we tuned the sparsification threshold $\epsilon \in \{10^{-2}, 10^{-3}, 10^{-4}\}$ for edge weight threshold applied to the diffusion matrix, the top-k

values $k \in \{32, 64, 128\}$ for number of neighbors retained per node after diffusion, heat diffusion time $t \in \{1, \ldots, 6\}$, and teleport probability $\alpha \in \{0.05, 0.10, 0.15, 0.20\}$.

- **GraphHeat** (Xu et al., 2019) and **ADC** (Zhao et al., 2021): We tuned the heat diffusion time in the range $t \in \{1, \ldots, 6\}$. For ADC, we also tuned the truncation order of the Taylor expansion $K \in \{5, 10, 15, 20\}$.
- **BrainGNN** (Li et al., 2021): We adjusted the pooling ratio for R-pool layer $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ and the number of ROI communities $K^{(l)} \in \{2, 4, 6, 8\}$.
- **BrainNetTF** (Kan et al., 2022): We tuned the number of transformer heads $\{2, 4, 8\}$ and the number of clustering centers $K \in \{2, 4, 8, 10, 20\}$ in the OCREAD readout module.
- **SCNN** (Yang et al., 2022): We implemented SCNN using node- and edge-level Hodge Laplacians ($\mathcal{L}_0$, $\mathcal{L}_1$) for simplicial convolution in our graph classification setting. We tuned the simplicial filter order $K \in \{2, 3, 4, 5\}$.
- **SaNN** (Gurugubelli & Chepuri, 2024): We implemented SaNN using node- and edge-level simplicial propagation based on $\mathcal{L}_0$ and $\mathcal{L}_1$. We tuned the propagation depth $K \in \{2, 3, 4, 5\}$.
- **ALTER** (Yu et al., 2024): We tuned the number of random walk hops $K \in \{2, 4, 8, 16, 32\}$ for the adaptive long-range encoding module. For the transformer encoder, we searched over the number of layers $L \in \{1, 2, 3, 4\}$ and the number of attention heads $M \in \{1, 2, 3, 4\}$.
- **ContrastPool** (Xu et al., 2024): We tuned the trade-off hyperparameters $\lambda_1 \in \{10, 1, 0.1\}$ and $\lambda_2 \in \{10^{-2}, 10^{-3}, 10^{-4}\}$ for the entropy regularization losses on the assignment matrix and the contrast graph, respectively. We also searched over the pooling ratio $\in \{0.3, 0.4, 0.5, 0.6\}$ and the number of pooling layers $L \in \{2, 3, 4\}$.

These settings follow the experimental protocol of AGT (Cho et al., 2024), which is a prior method using the same datasets. Full code and hyperparameter configurations will be released upon acceptance to ensure reproducibility.

F Additional Experimental Results

F.1 Quantitative Results for ABIDE dataset

We primarily select neurodegenerative disease datasets as our main experimental benchmark, since they involve subtle yet progressive changes in both regional brain activity and inter-regional connectivity across multiple disease stages. To additionally assess generality of our model beyond neurodegenerative settings, we further conduct experiments on the Autism Brain Imaging Data Exchange (ABIDE) (Heinsfeld et al., 2018) dataset.

ABIDE provides imaging and biomarkers for autism spectrum disorder (ASD). Functional brain networks are derived from fMRI on 116 regions based on the AAL atlas. Following the protocol in (Xu et al., 2023), we utilize Blood-Oxygen-Level-Dependent (BOLD) signals as node features and functional connectivity (correlation) between regions as edge features. The BOLD time-series preprocessing, correlation-based edge construction, and incidence-matrix construction follow the same procedure as described for PPMI in Sec. D.2. The dataset contains two diagnostic groups: Control and Autism. The demographic details of the ABIDE dataset are presented in Tab. 4 (Left).

The quantitative results on the ABIDE dataset are summarized in Tab. 4 (Right). Our method consistently outperforms all brain-specific baseline models across accuracy, precision, and recall, demonstrating its effectiveness on large-scale brain network classification beyond neurodegenerative settings.

Table 4: Left: Demographics of the ABIDE dataset. Right: Classification performance (Control vs. Autism) on the ABIDE dataset evaluated using 5-fold cross-validation across brain-specific models.

Category	**Control**	**Autism**
# of subjects	534	455
Gender (M / F)	441 / 93	399 / 56
Age (Mean ± Std)	16.9 ± 7.4	16.3 ± 7.6

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
BrainGNN	63.3 ± 2.4	63.6 ± 3.2	63.1 ± 4.6
BrainNetTF	70.3 ± 3.8	70.5 ± 5.5	70.0 ± 6.3
ALTER	76.8 ± 2.4	77.0 ± 2.1	76.5 ± 6.1
ContrastPool	69.6 ± 2.8	69.8 ± 4.3	69.3 ± 8.7
AGT	75.7 ± 1.7	76.0 ± 2.9	75.4 ± 5.1
Ours	**80.3±1.9**	**80.5±3.2**	**80.1±4.7**

F.2 Ablation study on expert number and top-k selection in ScaleX

We investigate the impact of varying the number of experts and top-k selection in ScaleX, as shown in Tab. 5 and Tab. 6. Experiments follow the same evaluation protocol as Tab. 1 on the ADNI dataset.

Increasing the number of node and edge experts enhances spectral representation by leveraging a more diverse set of wavelet scales. This improves the model's ability to capture both localized and global connectivity patterns across multiple resolutions. However, an excessive number of experts does not always yield better performance, as seen in Tab. 6, where accuracy saturates or slightly decreases when $P, Q = 7$. This is due to increased redundancy in spectral representation, where multiple experts learn overlapping frequency components rather than capturing distinct spectral features.

The top-k value, which controls the number of active experts per instance, also affects adaptability. A lower value may restrict scale diversity, whereas a higher value can introduce unnecessary complexity. An optimal top-k ensures the model focuses on the most relevant spectral scales, balancing efficiency and expressiveness.

Table 5: Ablation study on expert number and top-k selection in ScaleX for the ADNI-CT.

Number of Experts		**top-k**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
P	Q	k			
1	1	1	92.5 ± 1.1	92.0 ± 1.1	93.4 ± 1.0
3	3	1	93.3 ± 1.4	93.1 ± 1.6	93.5 ± 1.4
3	3	2	94.0 ± 1.2	93.8 ± 1.3	95.0 ± 1.0
3	3	3	93.9 ± 1.0	93.8 ± 0.9	95.0 ± 0.8
5	5	1	94.1 ± 1.8	93.7 ± 2.2	94.8 ± 1.2
5	5	2	**94.9 ± 1.3**	**95.4 ± 0.8**	**94.4 ± 1.7**
5	5	5	94.3 ± 1.3	94.0 ± 1.3	95.2 ± 1.1

Table 6: Ablation study on expert number and top-k selection in ScaleX for the ADNI-FDG.

Number of Experts		**top-k**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
P	Q	k			
1	1	1	92.8 ± 0.7	93.3 ± 1.3	93.1 ± 0.9
3	3	1	93.2 ± 1.1	91.7 ± 1.1	94.2 ± 1.0
3	3	2	95.5 ± 0.9	95.1 ± 1.5	95.6 ± 0.8
3	3	3	95.1 ± 0.9	94.9 ± 1.0	96.9 ± 1.2
4	5	2	96.2 ± 1.2	96.4 ± 1.3	96.1 ± 1.2
5	4	2	**96.4 ± 0.9**	**96.5 ± 1.1**	**97.1 ± 1.1**
5	5	1	94.4 ± 1.5	93.9 ± 1.3	94.6 ± 1.2
5	5	2	96.3 ± 0.7	96.2 ± 1.0	96.9 ± 0.9
5	5	3	96.1 ± 0.9	95.9 ± 0.7	97.0 ± 0.9
5	5	4	95.8 ± 1.1	95.6 ± 1.3	96.8 ± 1.1
5	5	5	95.5 ± 1.2	94.8 ± 0.9	95.8 ± 1.0
7	7	1	94.1 ± 1.7	93.7 ± 0.9	94.8 ± 1.3
7	7	2	96.2 ± 0.7	96.1 ± 1.1	97.0 ± 1.2
7	7	5	95.9 ± 1.3	96.1 ± 1.3	95.7 ± 0.8

F.3 Comparison of Routing Strategies

To further evaluate the effectiveness of the proposed signal-scale routing, we compare ScaleX with three simpler expert aggregation strategies while keeping the network architecture and training objective unchanged: (1) *Fixed Routing*, where a fixed expert is selected from the same expert pool for all graph elements; (2) *Random Routing*, where the same number (k) of experts are randomly selected; and (3) *Uniform Aggregation*, where all experts contribute equally without learned routing. Each variant is trained independently using the same experimental protocol and objective, with only the routing strategy changed. The results are summarized in Tab. 7.

Table 7: Comparison of different expert routing and aggregation strategies.

Routing Strategy	**ADNI-CT**			**PPMI**			**ABIDE**		
	Acc. (%)	**Pre. (%)**	**Rec. (%)**	**Acc. (%)**	**Pre. (%)**	**Rec. (%)**	**Acc. (%)**	**Pre. (%)**	**Rec. (%)**
Fixed Routing	90.7 ± 1.6	91.8 ± 1.3	87.9 ± 1.4	84.1 ± 2.5	77.3 ± 5.4	76.8 ± 2.3	75.8 ± 2.6	76.0 ± 3.1	75.5 ± 4.4
Random Routing	92.9 ± 2.7	92.6 ± 3.3	93.7 ± 2.9	84.9 ± 3.3	79.6 ± 5.6	80.2 ± 4.0	76.9 ± 4.9	77.1 ± 6.6	76.7 ± 7.2
Uniform Aggregation	92.7 ± 1.1	92.5 ± 1.5	93.3 ± 1.3	85.2 ± 2.3	78.8 ± 5.3	79.1 ± 2.7	77.0 ± 1.8	77.2 ± 3.0	76.8 ± 4.9
Learned Routing (ScaleX)	$\mathbf{94.9 \pm 1.3}$	$\mathbf{95.4 \pm 0.8}$	$\mathbf{94.4 \pm 1.7}$	$\mathbf{87.4 \pm 2.1}$	$\mathbf{82.1 \pm 3.4}$	$\mathbf{80.7 \pm 1.6}$	$\mathbf{80.3 \pm 1.9}$	$\mathbf{80.5 \pm 3.2}$	$\mathbf{80.1 \pm 4.7}$

Learned Routing achieves the highest mean accuracy across all three datasets. Since the network architecture and training objective are kept unchanged across the variants, the comparisons with Fixed Routing, Random Routing, and Uniform Aggregation highlight the benefit of adaptively routing graph elements to learned spectral experts rather than using non-adaptive expert assignments. These results support the effectiveness of learned routing for adaptive spectral filtering.

F.4 Ablation Study on Routing Balance

To evaluate the impact of routing balance regularization, we conduct an ablation study comparing model accuracy with and without L_{bal}. The results are summarized in Tab. 8. With L_{bal}, the model shows consistent improvements across all datasets, with an average accuracy gain of $\sim 1.93\%$p.

These improvements demonstrate the effectiveness of routing balance regularization in ScaleX. Without L_{bal}, expert assignments can become uneven, with a subset of experts receiving disproportionately high routing weights. By encouraging balanced expert utilization at the batch level, L_{bal} improves model stability and generalization.

Table 8: Comparison of accuracy with and without routing balance regularization.

L_{ce}	L_{bal}	ADNI-CT	ADNI-FDG	PPMI
✓		93.1 ± 1.9	94.1 ± 2.3	85.7 ± 2.9
✓	✓	$\mathbf{94.9 \pm 1.3}$ (+1.8)	$\mathbf{96.4 \pm 0.9}$ (+2.3)	$\mathbf{87.4 \pm 2.1}$ (+1.7)

F.5 Discussion on Model Complexity

Regarding the computational cost of spectral decomposition, which constitutes the major bottleneck of SSWT, a naive implementation would require separate eigen-decompositions for $\mathcal{L}_0$ and $\mathcal{L}_1$, incurring a total complexity of $\mathcal{O}(N^3 + M^3)$, where N and M denote the number of nodes and edges. However, as established in Proposition 1, our framework leverages a single SVD of the incidence matrix B_1 to simultaneously obtain the spectral bases for both nodes and edges. This effectively reduces the computational burden to that of one matrix factorization, $\mathcal{O}(\min(NM^2, N^2M))$. Furthermore, this operation is performed only once as a preprocessing step prior to training, which is a common practice in spectral methods. To evaluate the average runtime (over 10 epochs) and memory efficiency of our model, we compared it against several representative baselines, as summarized in Tab. 9. All experiments were conducted on an NVIDIA RTX 6000 Ada GPU, using a batch size that includes the entire ADNI-CT dataset.

Our model exhibits competitive efficiency compared to traditional GNN-based approaches (e.g., GAT) and achieves a substantial reduction in training time and memory consumption over recent transformer-based brain models (e.g., BrainNetTF). Notably, when comparing our sparse expert selection mechanism with a dense variant that utilizes all experts without top-k selection, the sparse version reduces runtime by 49% and memory usage by 24%.

Table 9: Comparison of average runtime (sec/epoch) and memory usage (MB) across models.

Model	**sec/epoch**	**memory (MB)**
GCN	0.0091	2,147
GAT	0.0915	10,657
BrainNetTF	0.1634	32,243
Ours w/o sparse MoE	0.1992	9,051
Ours	0.1012	6,861

F.6 Evaluation under Class Imbalance

To further assess performance under class imbalance, we report balanced accuracy, macro-F1, macro one-vs-rest AUROC, and class-wise recall on ADNI-CT. We compare our model with AGT, a strong baseline in the main comparison, along with SVM, GCN, and BrainGNN.

Table 10: Evaluation under class imbalance on ADNI-CT. Left: Additional classification metrics. Right: Class-wise recall (%).

Model	**Acc. (%)**	**Bal. Acc. (%)**	**Macro-F1 (%)**	**Macro OvR AUROC (%)**
SVM	82.54 ± 2.03	82.15 ± 2.33	83.73 ± 2.09	88.65 ± 1.44
GCN	59.86 ± 1.81	57.37 ± 2.32	58.32 ± 2.11	83.06 ± 1.72
BrainGNN	43.54 ± 1.56	37.42 ± 2.09	29.04 ± 3.69	65.04 ± 2.96
AGT	90.25 ± 1.77	89.57 ± 2.34	90.16 ± 2.37	92.19 ± 1.92
Ours	$\mathbf{94.89 \pm 1.33}$	$\mathbf{94.38 \pm 1.76}$	$\mathbf{94.77 \pm 1.34}$	$\mathbf{99.21 \pm 0.39}$

Model	**CN**	**SMC**	**EMCI**	**LMCI**	**AD**
SVM	84.1	74.6	83.3	77.8	91.0
GCN	59.6	43.6	69.3	46.1	68.1
BrainGNN	12.5	0.0	88.6	8.9	77.1
AGT	89.4	87.3	91.8	87.8	91.6
Ours	**95.0**	**91.1**	**96.1**	**93.9**	**95.8**

Our model achieves 94.38% balanced accuracy and 94.77% macro-F1, with class-wise recall ranging from 91.1% to 96.1% across the five diagnostic groups. These results indicate that the strong overall performance is maintained across individual classes rather than being dominated by the larger classes.

G DISCUSSIONS

Limitations. Although our model demonstrates robust performance across datasets when multiple experts are activated (i.e., $k \geq 2$), as shown in Tab. 5 and Tab. 6, the number of experts and the top-k value must still be carefully chosen to ensure optimal performance. Furthermore, the multi-resolution learning process consumes more memory resources than conventional models. However, due to the sparse activation of experts, our model avoids unnecessary computation, offering improved efficiency over dense expert model as in Tab. 9.

Broader Impact. This work presents a novel approach for analyzing brain connectomes, aiming to support the diagnosis of neurodegenerative diseases. The proposed method has the potential to positively impact society by enabling earlier and more accurate detection of conditions such as Alzheimer's and Parkinson's disease, leading to improved patient quality of life and reduced healthcare costs. However, overreliance on model predictions without clinical oversight may lead to misdiagnosis. To avoid this, the model should be used as a decision-support tool alongside expert interpretation within standard diagnostic workflows. These risks can be mitigated by incorporating expert interpretation when using the model in clinical workflows.