HAM-BGT: Harmonic-Aware Multimodal Brain Graph Transformer for Brain Disease Diagnosis

Heming Xu[1], Jingxi Feng[1], Dong Zhang[1], Zhiqiang Tian[1], and Shaoyi Du[1] ✉

Xi'an Jiaotong University, Xi'an 710049, China
xuheming@stu.xjtu.edu.cn {jingxifeng8,dongzhangcv}@gmail.com
{zhiqiangtian,dushaoyi}@xjtu.edu.cn

Abstract. Functional and structural brain networks characterize inter-regional functional coordination and anatomical connectivity, respectively, supporting brain disorder diagnosis and mechanistic studies. However, most existing multimodal brain network methods rely on simple feature concatenation, or treat functional and structural networks as independent graphs during fusion, overlooking their inherent coupling as revealed by connectome harmonics and thus failing to adequately capture cross-modal interactions for diagnosis. Specifically, eigenmodes of the structural connectome form a Fourier-like basis, and functional patterns can be described as energy distributed across these modes. To address these issues, we propose a Harmonic-Aware Multimodal Brain Graph Transformer (HAM-BGT) for deep joint modeling of multimodal brain networks. HAM-BGT introduces a Spectral Graph Filtering-Based Harmonic Encoding module to inject structural topological priors into functional connectivity representations. It then stacks multimodal brain graph Transformer layers to capture local-to-global dependencies. Each layer is followed by a Harmonic-Aware Filter that adaptively reorganizes regional features across frequency bands, enabling efficient extraction of harmonic-driven structure-function coupling patterns. Across multiple diagnostic tasks, HAM-BGT achieves state-of-the-art performance and identifies disease-relevant biomarkers, offering a harmonic-informed computational perspective to understand structure-function coupling in brain disorders.

Keywords: Cross-modal · Connectome harmonics · Transformer · Brain disorder diagnosis.

1 Introduction

Multimodal neuroimaging is critical for brain disorder diagnosis and mechanistic studies [18,4]. Functional brain networks capture the synchrony of neural activity across regions [20], whereas structural brain networks characterize anatomical connectivity [1]. Jointly analyzing both modalities helps reveal how structural wiring shapes and constrains functional coordination, offering complementary insights for brain disorder diagnosis. Recent studies reveal structure-function coupling in the connectome-harmonic domain, where spectral decomposition of

structural connectivity yields Fourier-like connectome harmonics that constrain functional patterns, and functional variations can be described as energy redistribution in the connectome-harmonic domain [2,3,14].

In recent years, graph neural networks (GNNs) have been widely used to exploit complementary information between multimodal brain networks for improved diagnosis [24,12,22], yet their local message passing limits capturing long-range dependencies between distant regions of interest (ROIs). Transformer-based models instead use self-attention to model global interactions and fuse multimodal networks globally [25]. However, these methods often overlook that functional and structural networks represent two views of the same set of ROIs with inherent connectome-harmonic coupling.

To address these issues, we propose a Harmonic-Aware Multimodal Brain Graph Transformer (HAM-BGT), explicitly modeling the structure-function coupling via connectome harmonics. HAM-BGT employs Spectral Graph Filtering-Based Harmonic Encoding to inject structural connectivity priors into functional representations, then uses multimodal brain graph Transformer layers to capture short- and long-range regional interactions. A Harmonic-Aware Filter after each layer enforces connectome-harmonic constraints across frequency bands during message passing, thereby enabling more accurate diagnosis of brain disorders. In summary, we make three main contributions:

1. We propose HAM-BGT, the first framework to explicitly model multimodal brain networks via connectome harmonics, offering a new avenue to investigate structure-function interactions.
2. We introduce Spectral Graph Filtering-Based Harmonic Encoding that jointly models structural and functional features at the node level, providing expressive initial representations for multimodal brain graph learning.
3. We propose a Harmonic-Aware Filter that enforces connectome-harmonic constraints during message passing, improving characterization of cross-modal coupling patterns and brain disorder identification.

2 Method

To address the lack of unified multimodal modeling via connectome harmonics, HAM-BGT employs three synergistic components that progressively inject, propagate, and refine harmonic constraints, as illustrated in Fig. 1:

(a) Harmonic-Aware Multimodal Regional Features Extraction: Injects structural priors into functional features via Spectral Graph Filtering-Based Harmonic Encoding (SGFBHE) to derive multimodal regional representations.
(b) Harmonic-Aware Multimodal Brain Graph Encoder: Stacks L layers, each combining a multimodal brain graph Transformer layer for capturing both short- and long-range dependencies among ROIs and a Harmonic-Aware Filter for enforcing connectome-harmonic constraints during message passing.
(c) Graph-Level Readout for Downstream Task: Aggregates node embeddings for downstream task.

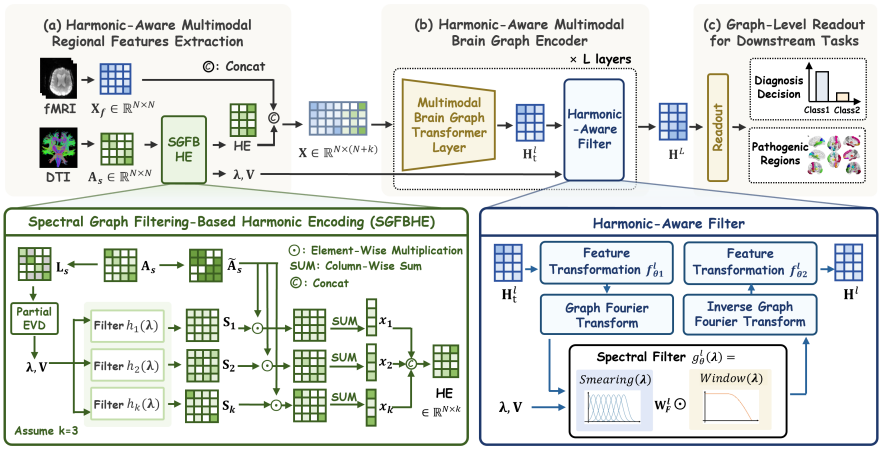

Fig. 1. The HAM-BGT framework. (a) Harmonic-Aware Multimodal Regional Features Extraction; (b) Harmonic-Aware Multimodal Brain Graph Encoder; (c) Graph-Level Readout for Downstream Tasks.

2.1 Harmonic-Aware Multimodal Regional Features Extraction

After standard preprocessing, functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) data are parcellated into N ROIs using the same atlas. From the fMRI data, we compute functional connectivity $\mathbf{X}_f \in \mathbb{R}^{N\times N}$ by measuring BOLD signal correlations between ROIs. From the DTI data, we compute structural connectivity $\mathbf{A}_s \in \mathbb{R}^{N\times N}$ by counting white matter fiber bundles between ROIs.

Phase synchrony of neural oscillations between ROIs in functional brain networks reflects functional activity and is shaped and constrained by connectome harmonics [2,3,21,19]. Motivated by this structure-function coupling, we augment the functional connectivity $\mathbf{X}_f$ with Spectral Graph Filtering-Based Harmonic Encoding (SGFBHE) to obtain multimodal regional features. First, following the definition of connectome harmonics, we construct the graph Laplacian of the structural connectivity $\mathbf{A}_s$ as $\mathbf{L}_s = \mathbf{I} - \mathbf{D}_s^{-\frac{1}{2}}\mathbf{A}_s\mathbf{D}_s^{-\frac{1}{2}}$, where $\mathbf{D}_s$ is the diagonal degree matrix. Second, to mitigate high-frequency noise, we perform a partial eigendecomposition of $\mathbf{L}_s$ and retain the first k eigenvalues $\boldsymbol{\lambda} \in \mathbb{R}^k$ and their corresponding eigenvectors $\mathbf{V} \in \mathbb{R}^{N\times k}$ to preserve low-frequency harmonics that characterize topology while suppressing noise:

$$(\boldsymbol{\lambda}, \mathbf{V}) = \text{Partial EVD}(\mathbf{L}_s, k). \tag{1}$$

Next, considering the sensitivity of eigenvectors to graph topology, we do not directly use $\mathbf{V}$ as the encoding. Instead, for each eigenvalue λ_i $(i = 1, 2, \ldots, k)$, we construct a radial basis filter centered at λ_i:

$$h_i(\boldsymbol{\lambda}) = \text{diag}\left(\text{softmax}(-\frac{(\lambda_i - \boldsymbol{\lambda}) \odot (\lambda_i - \boldsymbol{\lambda})}{\sigma^2})\right) \in \mathbb{R}^{k\times k}, \tag{2}$$

where $\odot$ denotes the Hadamard product. Intuitively, each $h_i(\boldsymbol{\lambda})$ acts as a band-pass filter centered at harmonic frequency λ_i, isolating the structural connectivity pattern associated with that particular spatial scale. Spectral filtering then yields a spatial domain representation of harmonic components around λ_i as $\mathbf{S}_i = \mathbf{V}h_i(\boldsymbol{\lambda})\mathbf{V}^\top \in \mathbb{R}^{N\times N}$. Moreover, since direct structural connections serve as the primary substrates for harmonic propagation [2], we binarize $\mathbf{A}_s$ to obtain $\tilde{\mathbf{A}}_s$ as an anatomical mask, and accumulate harmonic energy along verified pathways while suppressing spectral leakage to obtain harmonic-specific features:

$$\mathbf{x}_i = [\mathbf{S}_i \odot \tilde{\mathbf{A}}_s]\mathbf{1} \in \mathbb{R}^N, \tag{3}$$

where $\mathbf{1} \in \mathbb{R}^N$ is an all-ones vector and $\odot$ denotes the Hadamard product. Concatenating the features from all bands along the feature dimension yields the encoding $\mathrm{HE} = [\mathbf{x}_1\|\mathbf{x}_2\|\ldots\|\mathbf{x}_k] \in \mathbb{R}^{N\times k}$. Finally, we concatenate $\mathbf{X}_f$ with the harmonic encoding HE to obtain the multimodal regional features:

$$\mathbf{X} = [\mathbf{X}_f\|\mathrm{HE}] \in \mathbb{R}^{N\times(N+k)}, \tag{4}$$

where $\|$ denotes concatenation. $\mathbf{X}$ jointly characterizes functional features and harmonic-based structural features in the spatial domain and serves as the input to the subsequent harmonic-aware multimodal brain graph encoder.

2.2 Harmonic-Aware Multimodal Brain Graph Encoder

The multimodal regional features $\mathbf{X}$ are fed into an L-layer encoder. Each layer comprises a multimodal brain graph Transformer layer and a Harmonic-Aware Filter detailed below.

Multimodal Brain Graph Transformer Layer To capture long-range dependencies difficult for local GNN aggregation, we treat the brain network as fully connected and employ multi-head self-attention for message passing:

$$\mathbf{H}_t^l = (\|_{m=1}^M \mathbf{H}_t^{l,m})\mathbf{W}_O^l, \quad \mathbf{H}_t^{l,m} = \mathrm{softmax}(\frac{\mathbf{Q}^{l,m}(\mathbf{K}^{l,m})^\top}{\sqrt{d_k}})\mathbf{V}^{l,m}, \tag{5}$$

where $\mathbf{Q}^{l,m} = \mathbf{W}_Q^{l,m}\mathbf{H}^{l-1}$, $\mathbf{K}^{l,m} = \mathbf{W}_K^{l,m}\mathbf{H}^{l-1}$, and $\mathbf{V}^{l,m} = \mathbf{W}_V^{l,m}\mathbf{H}^{l-1}$ are the query, key, and value matrices of the m-th attention head, $m \in \{1, 2, \ldots, M\}$, M is the number of attention heads, $\mathbf{H}^{l-1}$ is the output from the previous layer with $\mathbf{H}^0 = \mathbf{X}$, and $\mathbf{W}_Q^{l,m}$, $\mathbf{W}_K^{l,m}$, $\mathbf{W}_V^{l,m}$, and $\mathbf{W}_O^l$ are learnable parameters, d_k is the first dimension of $\mathbf{W}_K^{l,m}$. The operator $\|$ denotes concatenation over heads. A position-wise feed-forward network and layer normalization are subsequently applied to obtain the final output of the Transformer layer $\mathbf{H}_t^l$.

Harmonic-Aware Filter To constrain the message passing through structure—function coupling induced by connectome harmonics, we apply adaptive spectral filtering to the output $\mathbf{H}_t^l$ of each multimodal brain graph Transformer layer:

$$\mathbf{H}^l = f_{\theta_2^l}\Big(\mathbf{V}\left(g_{\theta^l}(\boldsymbol{\lambda}) \odot \mathbf{V}^\top f_{\theta_1^l}(\mathbf{H}_t^l)\right)\Big), \tag{6}$$

where $\mathbf{V}^\top$ and $\mathbf{V}$ perform the graph Fourier transform (GFT) and its inverse (IGFT), $g_{\theta^l}(\boldsymbol{\lambda}) \in \mathbb{R}^{k\times d}$ is a spectral filter defined over the first k harmonics, and $\odot$ denotes the Hadamard product. Prior to the GFT, a gating mechanism $f_{\theta_1^l}$ provides lightweight node-wise adaptivity: $f_{\theta_1^l}(\mathbf{H}_t^l) = \mathbf{H}_t^l \odot \sigma'(\mathbf{H}_t^l\mathbf{W}_G^l + \mathbf{b}_G^l)$, where $\mathbf{W}_G^l \in \mathbb{R}^{d\times d}$, $\mathbf{b}_G^l \in \mathbb{R}^d$ are learnable parameters and σ' is the SiLU activation. The spectral filter is constructed as:

$$g_{\theta^l}(\boldsymbol{\lambda}) = \text{Smearing}(\boldsymbol{\lambda})\,\mathbf{W}_F^l \odot \text{Window}(\boldsymbol{\lambda}). \tag{7}$$

To ensure smooth yet expressive filter responses, each λ_i is first lifted into a high-dimensional representation via J Gaussian radial basis functions following [16]:

$$\text{Smearing}(\lambda_i) = \Big[\exp\Big(-\frac{(\lambda_i-\mu_1)^2}{2\sigma^2}\Big),\ \ldots,\ \exp\Big(-\frac{(\lambda_i-\mu_J)^2}{2\sigma^2}\Big)\Big] \in \mathbb{R}^{1\times J}, \tag{8}$$

where $\{\mu_j\}_{j=1}^J$ are Gaussian centers uniformly spaced over $[0,\ \lambda_{\text{cut}}]$, and J and λ_{cut} are hyperparameters. Stacking the expansions of all k eigenvalues yields $\text{Smearing}(\boldsymbol{\lambda}) \in \mathbb{R}^{k\times J}$, which is then mapped through a bias-free linear projection $\mathbf{W}_F^l \in \mathbb{R}^{J\times d}$ to produce the filter response. Since only the first k eigenmodes are retained, a Tukey window $\text{Window}(\boldsymbol{\lambda})$ is applied element-wise to smoothly attenuate high-frequency components, suppressing ringing artifacts from spectral truncation while preserving low-frequency structural information. After the IGFT, a linear projection with SiLU activation $f_{\theta_2^l}$ is applied to align the output distribution with the input of the subsequent Transformer layer.

2.3 Graph-level Readout and Training Objective

To obtain a graph-level representation for downstream tasks, we apply the Orthonormal Clustering Readout (OCREAD) [10] to aggregate the final node representations $\mathbf{H}^L \in \mathbb{R}^{N\times d}$ into a graph embedding $\mathbf{z} = \text{Readout}(\mathbf{H}^L)$. A two-layer MLP then predicts class probabilities: $\hat{\mathbf{y}} = \text{Softmax}\big(g(\mathbf{z})\big)$. The model is trained using cross-entropy loss.

3 Experiments

3.1 Experimental Settings

Datasets and Preprocessing Two public datasets were used, the Alzheimer's disease related dataset ADNI [9] and the autism spectrum disorder related dataset ABIDE [8]. The ADNI subset contains 468 subjects, including 94 subjects with Alzheimer's disease (AD), 174 subjects with mild cognitive impairment (MCI), and 200 normal controls (NCs). The ABIDE subset contains 95 subjects with both fMRI and DTI, including 42 subjects with autism spectrum disorder (ASD) and 53 NCs. The brain is parcellated into 90 ROIs according to the AAL atlas [15]. fMRI data are preprocessed using the DPARSF toolbox[1][23] to extract BOLD signals from each ROI, while DTI data are processed

[1] `http://rfmri.org/DPARSF`

Table 1. Comparative Experiments of Classification Tasks on Different Datasets.

Datasets (Tasks)		ABIDE (ASD vs. NC)				ADNI (AD vs. NC)			
Modal	Method	ACC(%)	SPE(%)	SEN(%)	AUC(%)	ACC(%)	SPE(%)	SEN(%)	AUC(%)
DTI	BrainGNN	$60.0_{\pm 4.1}$	$58.5_{\pm 8.1}$	$61.9_{\pm 4.0}$	$62.5_{\pm 1.6}$	$70.1_{\pm 2.8}$	$71.0_{\pm 3.4}$	$68.0_{\pm 6.9}$	$68.3_{\pm 1.0}$
	BrainIB	$61.1_{\pm 1.3}$	$60.4_{\pm 3.7}$	$61.9_{\pm 4.0}$	$61.4_{\pm 2.4}$	$70.4_{\pm 1.3}$	$70.0_{\pm 2.5}$	$71.3_{\pm 2.7}$	$70.8_{\pm 2.5}$
	BrainNetTF	$65.3_{\pm 1.5}$	$66.2_{\pm 3.5}$	$64.2_{\pm 2.3}$	$64.8_{\pm 4.5}$	$74.5_{\pm 2.6}$	$73.5_{\pm 4.2}$	$76.5_{\pm 5.0}$	$74.8_{\pm 0.8}$
	ALTER	$66.4_{\pm 4.2}$	$64.2_{\pm 3.6}$	$69.2_{\pm 5.6}$	$68.4_{\pm 3.7}$	$75.5_{\pm 4.8}$	$76.0_{\pm 6.8}$	$74.4_{\pm 2.6}$	$75.9_{\pm 1.7}$
fMRI	BrainGNN	$62.1_{\pm 2.0}$	$62.2_{\pm 2.0}$	$61.9_{\pm 4.0}$	$60.8_{\pm 2.8}$	$72.5_{\pm 1.7}$	$73.5_{\pm 2.2}$	$70.2_{\pm 5.8}$	$74.5_{\pm 2.7}$
	BrainIB	$63.1_{\pm 1.9}$	$62.2_{\pm 2.0}$	$64.2_{\pm 2.3}$	$64.3_{\pm 4.2}$	$73.5_{\pm 1.3}$	$73.0_{\pm 2.7}$	$74.5_{\pm 5.6}$	$73.0_{\pm 2.5}$
	BrainNetTF	$69.5_{\pm 2.1}$	$68.0_{\pm 4.1}$	$71.4_{\pm 6.5}$	$68.2_{\pm 2.6}$	$76.9_{\pm 1.8}$	$75.0_{\pm 3.1}$	$80.8_{\pm 3.1}$	$77.6_{\pm 3.0}$
	ALTER	$71.6_{\pm 1.6}$	$71.6_{\pm 1.5}$	$71.7_{\pm 4.6}$	$70.9_{\pm 4.1}$	$78.2_{\pm 1.6}$	$78.5_{\pm 3.8}$	$77.6_{\pm 6.1}$	$76.2_{\pm 2.1}$
Both	MS-Inter-GCN	$68.4_{\pm 1.7}$	$71.6_{\pm 1.5}$	$64.2_{\pm 2.3}$	$67.3_{\pm 3.8}$	$77.2_{\pm 0.9}$	$78.0_{\pm 2.7}$	$75.5_{\pm 3.2}$	$76.4_{\pm 1.6}$
	BrainNN	$69.4_{\pm 3.4}$	$73.5_{\pm 4.9}$	$64.2_{\pm 2.3}$	$68.4_{\pm 1.9}$	$75.9_{\pm 2.4}$	$76.5_{\pm 2.9}$	$74.5_{\pm 1.8}$	$72.8_{\pm 4.0}$
	RH-BrainFS	$71.6_{\pm 1.6}$	$71.6_{\pm 1.5}$	$71.7_{\pm 4.6}$	$71.5_{\pm 3.4}$	$76.9_{\pm 1.8}$	$76.5_{\pm 2.9}$	$77.7_{\pm 4.1}$	$78.6_{\pm 1.5}$
	Cross-GNN	$73.6_{\pm 1.4}$	$71.6_{\pm 1.5}$	$76.1_{\pm 1.5}$	$72.3_{\pm 2.1}$	$77.9_{\pm 4.0}$	$78.5_{\pm 3.4}$	$76.5_{\pm 5.5}$	$77.2_{\pm 2.1}$
	MME-GCN	$73.8_{\pm 3.9}$	$75.6_{\pm 4.0}$	$71.7_{\pm 4.6}$	$74.1_{\pm 2.4}$	$79.6_{\pm 2.3}$	$78.0_{\pm 4.5}$	$82.9_{\pm 2.9}$	$79.0_{\pm 3.1}$
	Ours	$\mathbf{78.9}_{\pm 1.1}$	$\mathbf{81.1}_{\pm 1.0}$	$\mathbf{76.1}_{\pm 1.5}$	$\mathbf{81.4}_{\pm 2.0}$	$\mathbf{84.7}_{\pm 1.3}$	$\mathbf{85.5}_{\pm 2.1}$	$\mathbf{82.9}_{\pm 4.7}$	$\mathbf{84.2}_{\pm 1.4}$

Datasets (Tasks)		ADNI(AD vs. MCI)				ADNI(MCI vs. NC)			
Modal	Method	ACC(%)	SPE(%)	SEN(%)	AUC(%)	ACC(%)	SPE(%)	SEN(%)	AUC(%)
DTI	BrainGNN	$63.8_{\pm 2.3}$	$64.4_{\pm 2.2}$	$62.8_{\pm 3.1}$	$63.6_{\pm 1.9}$	$60.7_{\pm 1.5}$	$62.0_{\pm 1.1}$	$59.2_{\pm 2.4}$	$60.0_{\pm 2.7}$
	BrainIB	$64.6_{\pm 1.8}$	$65.0_{\pm 2.1}$	$63.9_{\pm 1.6}$	$63.6_{\pm 2.3}$	$60.9_{\pm 4.1}$	$62.0_{\pm 6.0}$	$59.7_{\pm 5.2}$	$61.7_{\pm 3.8}$
	BrainNetTF	$67.2_{\pm 1.7}$	$65.5_{\pm 1.7}$	$70.2_{\pm 2.5}$	$67.3_{\pm 2.2}$	$62.0_{\pm 3.2}$	$62.0_{\pm 4.8}$	$62.1_{\pm 4.9}$	$60.5_{\pm 2.0}$
	ALTER	$69.0_{\pm 1.5}$	$70.1_{\pm 1.5}$	$67.0_{\pm 2.3}$	$66.9_{\pm 2.3}$	$62.3_{\pm 1.6}$	$63.5_{\pm 2.2}$	$60.9_{\pm 5.1}$	$62.0_{\pm 3.3}$
fMRI	BrainGNN	$66.8_{\pm 1.9}$	$67.8_{\pm 2.1}$	$64.9_{\pm 6.2}$	$65.1_{\pm 3.1}$	$61.2_{\pm 3.3}$	$62.5_{\pm 4.7}$	$59.8_{\pm 5.6}$	$60.6_{\pm 1.7}$
	BrainIB	$68.3_{\pm 2.3}$	$69.0_{\pm 2.5}$	$67.0_{\pm 2.3}$	$69.1_{\pm 2.0}$	$61.0_{\pm 1.3}$	$62.0_{\pm 1.1}$	$59.8_{\pm 2.1}$	$61.2_{\pm 2.2}$
	BrainNetTF	$70.5_{\pm 1.7}$	$70.1_{\pm 4.0}$	$71.3_{\pm 2.7}$	$71.4_{\pm 1.7}$	$63.6_{\pm 0.6}$	$65.0_{\pm 0.0}$	$62.1_{\pm 1.2}$	$64.1_{\pm 0.4}$
	ALTER	$71.3_{\pm 3.1}$	$70.1_{\pm 3.0}$	$73.5_{\pm 3.3}$	$72.5_{\pm 1.9}$	$64.2_{\pm 1.9}$	$63.5_{\pm 3.8}$	$64.9_{\pm 2.7}$	$64.4_{\pm 3.9}$
Both	MS-Inter-GCN	$71.6_{\pm 0.8}$	$71.3_{\pm 2.1}$	$72.3_{\pm 2.3}$	$72.5_{\pm 3.2}$	$66.3_{\pm 2.6}$	$65.0_{\pm 4.7}$	$67.8_{\pm 3.9}$	$68.4_{\pm 2.0}$
	BrainNN	$70.6_{\pm 2.7}$	$70.7_{\pm 3.6}$	$70.3_{\pm 6.7}$	$71.4_{\pm 2.7}$	$65.8_{\pm 2.3}$	$66.5_{\pm 2.2}$	$64.9_{\pm 2.4}$	$65.0_{\pm 1.7}$
	RH-BrainFS	$71.7_{\pm 2.0}$	$73.0_{\pm 2.3}$	$69.2_{\pm 1.7}$	$71.1_{\pm 1.2}$	$66.8_{\pm 2.2}$	$68.5_{\pm 2.9}$	$64.9_{\pm 4.6}$	$68.0_{\pm 1.8}$
	Cross-GNN	$72.8_{\pm 1.8}$	$72.4_{\pm 3.0}$	$73.4_{\pm 3.8}$	$\mathbf{75.3}_{\pm 1.0}$	$67.1_{\pm 2.1}$	$65.5_{\pm 2.1}$	$69.0_{\pm 2.2}$	$66.3_{\pm 2.4}$
	MME-GCN	$72.8_{\pm 0.8}$	$71.8_{\pm 3.2}$	$74.5_{\pm 4.1}$	$72.7_{\pm 1.9}$	$66.3_{\pm 2.3}$	$66.5_{\pm 3.4}$	$66.1_{\pm 3.6}$	$67.2_{\pm 1.9}$
	Ours	$\mathbf{75.7}_{\pm 1.9}$	$\mathbf{75.3}_{\pm 1.7}$	$\mathbf{76.6}_{\pm 2.7}$	$75.1_{\pm 1.7}$	$\mathbf{70.9}_{\pm 1.8}$	$\mathbf{70.0}_{\pm 3.5}$	$\mathbf{71.8}_{\pm 2.5}$	$\mathbf{72.4}_{\pm 3.5}$

using PANDA [5], where deterministic fiber tracking is applied to compute the number of white-matter fibers between ROIs.

Implementation details and evaluation metrics All experiments are implemented in PyTorch and conducted on an NVIDIA RTX 3090 GPU. We adopt 5-fold cross-validation with a class-balanced 4:1 train/test split. The model is trained for 300 epochs using the AdamW optimizer [13] with a learning rate of 1×10^{-4}. We use k=40 eigencomponents, M=2 attention heads, J=50 Gaussian radial basis functions, and set λ_{cut} to 1.1 times the largest eigenvalue λ_k. Evaluation metrics include accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under the ROC curve (AUC).

3.2 Comparison Experiment

Baseline methods include unimodal approaches BrainGNN [11], BrainIB [27], BrainNetTF [10], and ALTER [26], and multimodal approaches BrainNN [29],

MS-Inter-GCN [22], RH-BrainFS [25], Cross-GNN [24], and MME-GCN [12]. As shown in Table 1, the proposed method achieves the highest accuracy on all datasets, demonstrating both effectiveness and robustness. Specifically, on the ASD vs. NC task, HAM-BGT achieves an accuracy of 78.9%, outperforming the best unimodal methods (DTI: ACC=66.4%, fMRI: ACC=71.6%) as well as the best multimodal baseline (ACC=73.8%). Similar improvements are observed in other tasks. These gains are primarily attributed to the ability of the proposed Transformer-based framework to effectively capture both short- and long-range dependencies in multimodal brain networks, while better characterizing cross-modal interactions under connectome-harmonic constraints.

3.3 Ablation Studies

Ablation studies are conducted on the ASD vs. NC and AD vs. NC tasks to evaluate the contribution of each component in HAM-BGT, with results shown in Fig. 2. For the SGFBHE module, removing the encoding and using only unimodal functional connectivity as input, or replacing it with a Laplacian encoding, leads to performance drops exceeding 5%. This indicates that the SGFBHE module enables stable early fusion of multimodal features. Removing the Harmonic-Aware Filter also results in degraded performance, which suggests that this module, by adaptively reallocating spectral bands of regional features after the Transformer-based message passing, can effectively capture multimodal brain network interactions under connectome-harmonic constraints and further enhance diagnostic performance.

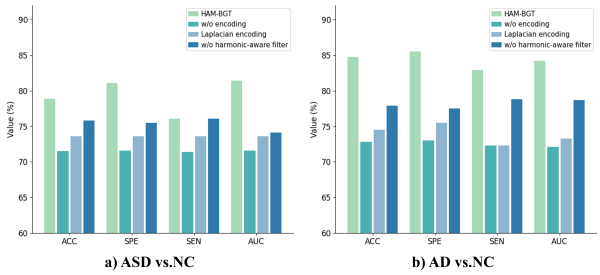

Fig. 2. Ablation Study Results on Different Datasets.

3.4 Interpretation Analysis

Discriminative ROIs Analysis To assess the impact of different ROIs in classification, we perform a t-test to measure node importance and select the top 10 most discriminative ROIs for each task, as shown in Fig. 3(1). For the ASD vs. NC task, the most discriminative ROIs include the precuneus (PCUN) and

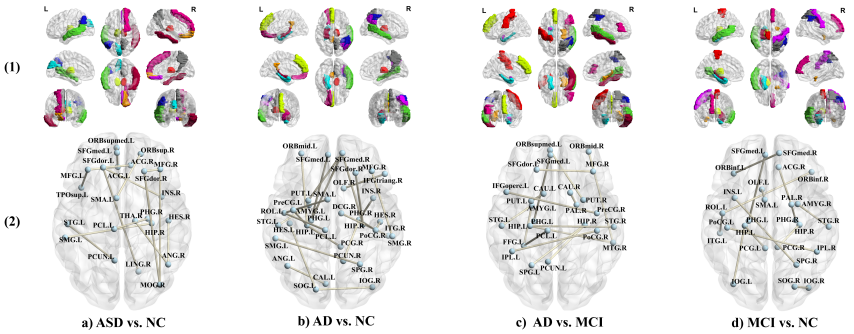

Fig. 3. Visualization of (1) discriminative ROIs and (2) discriminative connections.

rectus gyrus (REC). Prior studies have reported that these ROIs are involved in social cognition and self-referential processing, and their activation is reduced when individuals with ASD infer their own or others' mental states [17], which is consistent with our findings. For the AD vs. NC task, prominent ROIs include the angular gyrus (ANG) and hippocampus (HIP), reflecting characteristic impairments in memory and semantic processing in Alzheimer's disease [28,7].

Discriminative Connections Analysis We further evaluate the attention maps from the multimodal brain graph Transformer layers using a standard t-test and display significantly different connections (p-value < 0.01) in Fig. 3(2). By counting the important connections associated with each ROI, we observe that in the ASD vs. NC task these connections and ROIs are mainly distributed in the frontal lobe with dense connections involving the superior frontal gyrus (SFG) and the middle frontal gyrus (MFG) [6]. In the AD vs. NC task, important connections and ROIs mainly involve SFG, consistent with previous findings [28].

4 Conclusion

This work proposes a Harmonic-Aware Multimodal Brain Graph Transformer (HAM-BGT), which leverages the complementary information of functional and structural connectivity from a connectome-harmonic perspective for brain disorder diagnosis. HAM-BGT comprises two core modules: a Spectral Graph Filtering-Based Harmonic Encoding module that jointly models structural and functional features from a harmonic viewpoint to obtain more expressive multimodal regional representations; and a Harmonic-Aware Filter module that introduces connectome-harmonic constraints into message passing to effectively capture interactions across multimodal brain networks. Experiments on ABIDE and ADNI demonstrate that the proposed method outperforms state-of-the-art approaches and identifies meaningful brain network biomarkers.

Acknowledgments. This work was supported by the National Key Research and Development Program of China under Grant No. 2025ZD0217300, Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China under Grant No. JYB2025XDXM504, the National Natural Science Foundation of China under Grant Nos. 62501462 and 62125305, Guangdong Major Project of Basic and Applied Basic Research under Grant No. 2023B0303000009 and the Postdoctoral Research Funding Project of Shaanxi Province under Grant No. 2025BSHSDZZ023.

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

References

1. Assaf, Y., Pasternak, O.: Diffusion tensor imaging (dti)-based white matter mapping in brain research: a review. Journal of molecular neuroscience 34(1), 51–61 (2008)
2. Atasoy, S., Donnelly, I., Pearson, J.: Human brain networks function in connectome-specific harmonic waves. Nature communications 7(1), 10340 (2016)
3. Atasoy, S., Roseman, L., Kaelen, M., Kringelbach, M.L., Deco, G., Carhart-Harris, R.L.: Connectome-harmonic decomposition of human brain activity reveals dynamical repertoire re-organization under lsd. Scientific reports 7(1), 17661 (2017)
4. Calhoun, V.D., Sui, J.: Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness. Biological psychiatry: cognitive neuroscience and neuroimaging 1(3), 230–244 (2016)
5. Cui, Z., Zhong, S., Xu, P., He, Y., Gong, G.: Panda: a pipeline toolbox for analyzing brain diffusion images. Frontiers in human neuroscience 7, 42 (2013)
6. Dapretto, M., Davies, M.S., Pfeifer, J.H., Scott, A.A., Sigman, M., Bookheimer, S.Y., Iacoboni, M.: Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. Nature neuroscience 9(1), 28–30 (2006)
7. Drzezga, A., Becker, J.A., Van Dijk, K.R., Sreenivasan, A., Talukdar, T., Sullivan, C., Schultz, A.P., Sepulcre, J., Putcha, D., Greve, D., et al.: Neuronal dysfunction and disconnection of cortical hubs in non-demented subjects with elevated amyloid burden. Brain 134(6), 1635–1646 (2011)
8. Heinsfeld, A.S., Franco, A.R., Craddock, R.C., Buchweitz, A., Meneguzzi, F.: Identification of autism spectrum disorder using deep learning and the abide dataset. NeuroImage: Clinical 17, 16–23 (2018)
9. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 27(4), 685–691 (2008)
10. Kan, X., Dai, W., Cui, H., Zhang, Z., Guo, Y., Yang, C.: Brain network transformer. Advances in Neural Information Processing Systems 35, 25586–25599 (2022)
11. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L.H., Ventola, P., Duncan, J.S.: Braingnn: Interpretable brain graph neural network for fmri analysis. Medical Image Analysis 74, 102233 (2021)
12. Liu, L., Wang, Y.P., Wang, Y., Zhang, P., Xiong, S.: An enhanced multi-modal brain graph network for classifying neuropsychiatric disorders. Medical image analysis 81, 102550 (2022)

13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
14. Luppi, A.I., Vohryzek, J., Kringelbach, M.L., Mediano, P.A., Craig, M.M., Adapa, R., Carhart-Harris, R.L., Roseman, L., Pappas, I., Peattie, A.R., et al.: Distributed harmonic patterns of structure-function dependence orchestrate human consciousness. Communications biology 6(1), 117 (2023)
15. Rolls, E.T., Huang, C.C., Lin, C.P., Feng, J., Joliot, M.: Automated anatomical labelling atlas 3. Neuroimage 206, 116189 (2020)
16. Schütt, K., Kindermans, P.J., Sauceda Felix, H.E., Chmiela, S., Tkatchenko, A., Müller, K.R.: Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. Advances in neural information processing systems 30 (2017)
17. Silk, T.J., Rinehart, N., Bradshaw, J.L., Tonge, B., Egan, G., O'Boyle, M.W., Cunnington, R.: Visuospatial processing and the function of prefrontal-parietal networks in autism spectrum disorders: a functional mri study. American Journal of Psychiatry 163(8), 1440–1443 (2006)
18. Sui, J., Adali, T., Yu, Q., Chen, J., Calhoun, V.D.: A review of multivariate methods for multimodal fusion of brain imaging data. Journal of neuroscience methods 204(1), 68–81 (2012)
19. Uhlhaas, P.J., Singer, W.: Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. Neuron 52(1), 155–168 (2006)
20. Van Den Heuvel, M.P., Pol, H.E.H.: Exploring the brain network: a review on resting-state fmri functional connectivity. European Neuropsychopharmacology 20(8), 519–534 (2010)
21. Varela, F., Lachaux, J.P., Rodriguez, E., Martinerie, J.: The brainweb: phase synchronization and large-scale integration. Nature reviews neuroscience 2(4), 229–239 (2001)
22. Xia, J., Chan, Y.H., Girish, D., Rajapakse, J.C.: Interpretable modality-specific and interactive graph convolutional network on brain functional and structural connectomes. Medical Image Analysis 102, 103509 (2025)
23. Yan, C., Zang, Y.: Dparsf: a matlab toolbox for 'pipeline' data analysis of resting-state fmri. Frontiers in systems neuroscience 4, 1377 (2010)
24. Yang, Y., Ye, C., Guo, X., Wu, T., Xiang, Y., Ma, T.: Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning. IEEE Transactions on Medical Imaging 43(1), 108–121 (2023)
25. Ye, H., Zheng, Y., Li, Y., Zhang, K., Kong, Y., Yuan, Y.: Rh-brainfs: regional heterogeneous multimodal brain networks fusion strategy. Advances in Neural Information Processing Systems 36, 59286–59303 (2023)
26. Yu, S., Jin, S., Li, M., Sarwar, T., Xia, F.: Long-range brain graph transformer. Advances in Neural Information Processing Systems 37, 24472–24495 (2024)
27. Zheng, K., Yu, S., Li, B., Jenssen, R., Chen, B.: Brainib: Interpretable brain network-based psychiatric diagnosis with graph information bottleneck. IEEE Transactions on Neural Networks and Learning Systems (2024)
28. Zhu, H., Zhou, P., Alcauter, S., Chen, Y., Cao, H., Tian, M., Ming, D., Qi, H., Wang, X., Zhao, X., et al.: Changes of intranetwork and internetwork functional connectivity in alzheimer's disease and mild cognitive impairment. Journal of neural engineering 13(4), 046008 (2016)
29. Zhu, Y., Cui, H., He, L., Sun, L., Yang, C.: Joint embedding of structural and functional brain networks with graph neural networks for mental illness diagnosis. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 272–276. IEEE (2022)