DHGFormer: Dynamic Hierarchical Graph Transformer for Disorder Brain Disease Diagnosis

Rundong Xue[1], Hao Hu[1], Zeyu Zhang[2], Xiangmin Han[2(✉)], Juan Wang[3], Yue Gao[2], and Shaoyi Du[1(✉)]

[1] State Key Laboratory of Human-Machine Hybrid Augmented Intelligence, National Engineering Research Center for Visual Information and Applications, and Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, 710049, China
{xuerundong, huhao}@stu.xjtu.edu.cn, dushaoyi@xjtu.edu.cn
[2] Tsinghua University, 100084, China
steve.zeyu.zhang@outlook.com,{hanxiangmin, gaoyue}@tsinghua.edu.cn
[3] Department of Ultrasound, the Second Affiliated Hospital of Xi'an Jiaotong University, 710049, China

Abstract. The functional brain network exhibits a hierarchical characterized organization, balancing localized specialization with global integration through multi-scale hierarchical connectivity. While graph-based methods have advanced brain network analysis, conventional graph neural networks (GNNs) face interpretational limitations when modeling functional connectivity (FC) that encodes excitatory/inhibitory distinctions, often resorting to oversimplified edge weight transformations. Existing methods usually inadequately represent the brain's hierarchical organization, potentially missing critical information about multi-scale feature interactions. To address these limitations, we propose a novel brain network generation and analysis approach–Dynamic Hierarchical Graph Transformer (DHGFormer). Specifically, our method introduces an FC-inspired dynamic attention mechanism that adaptively encodes brain excitatory/inhibitory connectivity patterns into transformer-based representations, enabling dynamic adjustment of the functional brain network. Furthermore, we design hierarchical GNNs that consider prior functional subnetwork knowledge to capture intra-subnetwork homogeneity and inter-subnetwork heterogeneity, thereby enhancing GNN performance in brain disease diagnosis tasks. Extensive experiments on the ABIDE and ADNI datasets demonstrate that DHGFormer consistently outperforms state-of-the-art methods in diagnosing neurological disorders. The code is available at https://github.com/iMoonLab/DHGFormer.

Keywords: Brain Network · Graph Transformer · Brain Disease.

1 Introduction

The human brain operates as a complex network of interacting neural systems, where both localized functional specialization and distributed connectivity patterns underlie cognitive processes and behavioral outcomes [1,2,3]. This net-

work paradigm has revolutionized neuroimaging research, providing critical insights into normative brain organization and pathological alterations in neurological/psychiatric disorders[4,5].

In general, due to the correlation between different brain regions, the functional brain network can be naturally structured as a graph, where nodes represent regions of interest (ROI) and edges represent the FC. As a result, graph-based methods have become widely applied in brain network analysis. Although the correlation score-based (*e.g.* Pearson correlation coefficient, PCC) brain network generation mechanism is widely used in existing literature [6,7,8,9,10], it brings two limitations: GNNs are incompatible with positive and negative weighted edges inherent to the FC matrix, forcing ad hoc solutions like absolute value transformations that discard neurobiologically meaningful inhibitory/excitatory distinctions; and current graph construction methods either rely on oversimplified linear correlations or biologically ungrounded random graphs, failing to reconcile data-driven learning with established biological insights [11,12,13].

Furthermore, emerging neuroimaging evidence reveals that the brain's functional architecture exhibits an inherent hierarchical organization, comprising specialized subnetworks that coordinate both intra-modular processing and inter-modular communication [14]. This multi-scale structure suggests that functional subnetworks maintain distinct connectivity profiles: dense homogeneous connections within modules facilitate localized computation, while sparser heterogeneous links between modules enable global integration [15]. Current analytical frameworks frequently overlook this hierarchical structure or fail to integrate it comprehensively [16,17], potentially missing critical information about cross-scale interactions that could enhance disease characterization and mechanistic understanding.

To address the above problems, we present the Dynamic Hierarchical Graph Transformer (DHGFormer), a novel framework that synergizes dynamic graph adaptation with hierarchical representation learning for enhanced brain network analysis. Specifically, we introduce a dynamic brain transformer inspired by FC, which dynamically adjusts the connection weights between brain regions to learn task-aware connectivity patterns. Furthermore, we integrate a hierarchical GNN structure, considering prior the relationships between and within brain functional subnetworks, to model the brain's multi-scale architecture, thus providing a more comprehensive and flexible representation of brain communication propagation. The main contributions of this work are as follows:

1) We propose an FC-inspired dynamic brain transformer that enables adaptive encoding of brain excitatory/inhibitory connectivity patterns. By leveraging FC as an attention guiding factor to regulate the distribution of attention, this mechanism prioritizes biologically meaningful connections between brain regions during dynamic graph adjustment, thus optimizing the source of task-aware information.
2) By integrating prior knowledge about functional subnetworks, we design a brain-hierarchical GNN framework that separately models intra-subnetwork homogeneity and inter-subnetwork heterogeneity. Our cross-scale message-

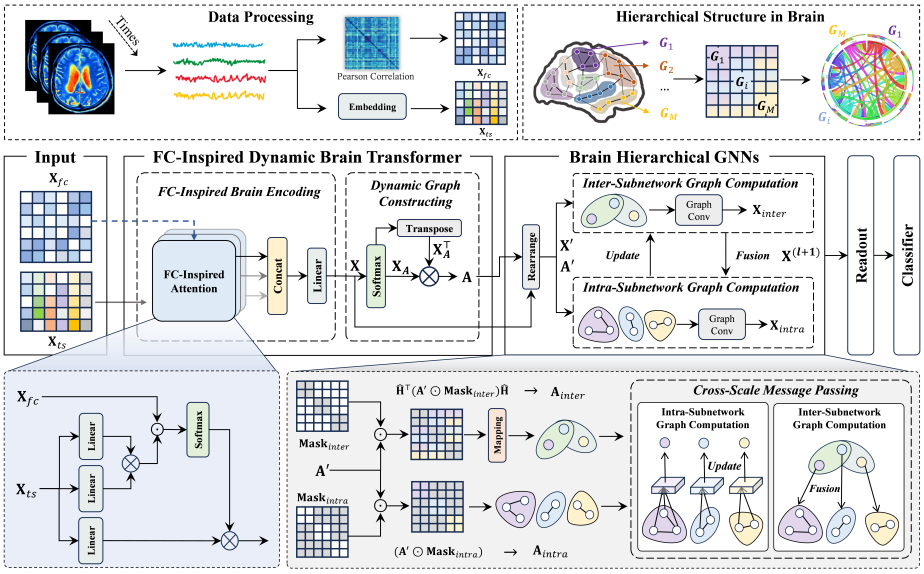


Fig. 1. The framework of the proposed DHGFormer.

passing mechanism enables information fusion and updating from the microscopic brain regions to the macroscopic functional subnetworks.

3) Experimental evaluations on two commonly used brain disease datasets demonstrate that our method outperforms all baseline methods. Additionally, the visualization of multi-scale brain region associations reveals the functional connectivity patterns at multiple levels, which is beneficial for uncovering brain pathological mechanisms.

2 Method

2.1 FC-Inspired Dynamic Brain Transformer

FC-Inspired Brain Encoding. Each scanned brain space is parcellated into N regions of interest (ROIs). We extract the BOLD time series $\mathbf{X}_{ts} \in \mathbb{R}^{N \times T}$ based on a given atlas, employing these to calculate the PCC matrix $\mathbf{X}_{fc} \in \mathbb{R}^{N \times N}$ to quantify the functional connectivity strength between each pair of ROIs. Existing functional brain networks are mainly constructed by calculating the PCC matrix. However, GNNs are incompatible with functional brain networks that have both positive and negative weighted edges, leading to ad-hoc solutions like taking absolute values, which ignore the distinct roles of positive/negative correlations (e.g., synergy vs. inhibition). Furthermore, predefined networks cannot adaptively optimize connectivity patterns for downstream tasks.

To solve the above problems, this paper proposes an FC-inspired downstream task-aware brain network generation method. Unlike [11], the proposed method

ingeniously applies FC guidance information $\mathbf{X}_{fc}$ to the attention mechanism, breaking through the inherent pattern of relying exclusively on the features learned by the model. This innovative operation guides the model to prioritize biologically meaningful connectivity during dynamic graph construction and optimize information processing at the source, thus achieving task-aware optimization while maintaining interpretability.

Specifically, the FC-inspired brain encoder is employed to extract potential functional correlations between brain regions from the time series $\mathbf{X}_{ts}$. Given the input $\mathbf{X}_{ts}$ and $\mathbf{X}_{fc}$, the encoder produces the brain region features in a global perspective. This process can be formulated as: $\mathbf{X} = \text{Encoder}(\mathbf{X}_{ts}, \mathbf{X}_{fc})$. For the FC-inspired brain encoder, we adopt a BNT-based encoder [18] as the backbone. By feeding the $\mathbf{X}_{ts}$ and $\mathbf{X}_{fc}$ into the encoder, the correlation representations $\mathbf{X}$ among ROIs are captured by multi-head FC-inspired attention mechanism:

$$\mathbf{X} = W_o(\|_{m=1}^{M} \text{head}_m) \text{ and } \text{head}_m = \text{softmax}\Big(\frac{Q^m(K^m)^\top}{\sqrt{d_k^m}} \odot \mathbf{X}_{fc}\Big)V^m, \quad (1)$$

where Q^m, K^m, V^m are the query, key, and value matrices for the m-th attention head, M is the number of attention heads, and W_o are learnable parameters. The Hadamard product of X, used as a prior attention weight distribution, suppresses irrelevant ROI connections and strengthens functionally significant ones.

Dynamic Graph Constructing. Based on the encoded time series feature $\mathbf{X}$, a learnable and interpretable method for generating functional brain networks is employed, which can be formulated as:

$$\mathbf{A} = \mathbf{X}_A\mathbf{X}_A^\top, \text{ where } \mathbf{X}_A = \text{softmax}(\mathbf{X}). \quad (2)$$

The generation of non-negative symmetric matrices through the operation of self-multiplication of transpose operations, in conjunction with the highlighting of significant ROI connections by generating skewed positive edge weights through softmax operations, is compatible with GNNs while adaptively integrating the excitatory/inhibitory brain connectivity patterns. The constructed functional graph structure can be dynamically optimized based on downstream tasks.

2.2 Brain Hierarchical GNNs

The human brain connectome is a hierarchical structure, and ROIs within the same functional subnetwork have greater functional similarities compared to functional correlations between functional subnetworks. Therefore, a brain hierarchical GNN is introduced to model this hierarchical structure and efficiently utilize the prior functional partitioning knowledge to learn local correlations within and long-range dependencies across subnetworks, which compensates for the shortcomings of long-range perception in conventional GNNs.

Intra-Subnetwork Graph Computation. To perform the intra-subnetwork graph computation, we first consider the following elements. Given M functional subnetworks $(G_1, \ldots, G_M)$ and the membership of ROIs in the Yeo [19],

we rearrange the rows and columns of the node feature matrix $\mathbf{X}$ and the adjacency matrix $\mathbf{A}$ of the output of Dynamic Brain Transformer, resulting in $\mathbf{X}'$ and $\mathbf{A}'$. The intra-subnetwork connectivity mask $\mathbf{Mask}_{intra} \in \mathbb{R}^{N\times N}$ is constructed as a binary matrix identifying ROI pairs within identical subnetworks: $\mathbf{Mask}_{intra}^{i,j} = 1 \Leftrightarrow v_i, v_j \in G_k (k = 1, \ldots, M)$. This mask selectively preserves functional associations between region pairs v_i and v_j that co-reside in any subnetwork G_k, while eliminating inter-subnetwork connectivity. Through Eq. 3, we extract intra-subnetwork connectivity patterns that reflect localized neural coordination, forming the basis for subsequent group-level analyses. With the node features $\mathbf{X}'$ and the intra-subnetwork $\mathbf{A}_{intra}$, we apply GNNs to update the intra-subnetwork feature representations. Compared to [16], which employs M parallel encoders for each subnetwork, our intra-subnetwork computation achieves efficient feature extraction via a single graph convolution, reducing parameters while improving accuracy as shown in Tab 1.

$$\mathbf{A}_{intra} = \mathbf{A}' \odot \mathbf{Mask}_{intra}, \tag{3}$$

$$\mathbf{X}_{intra}^{(l+1)} = \sigma(\mathbf{A}_{intra}\mathbf{X}^{(l)}\mathbf{W}^{(l)}), \text{ where } \mathbf{X}^{(0)} = \mathbf{X}'. \tag{4}$$

Inter-Subnetwork Graph Computation. To establish functional associations among subnetworks while preserving their global topological characteristics, we develop a cross-scale feature mapping method that enables projecting ROI-level features to group-level representations and deriving a group-level adjacency matrix that quantifies inter-subnetwork connectivity patterns.

$$\mathbf{Mask}_{inter} = \mathbf{I} - \mathbf{Mask}_{intra}, \tag{5}$$

$$\mathbf{A}_{inter} = \hat{\mathbf{H}}^{\top}(\mathbf{A}' \odot \mathbf{Mask}_{inter})\hat{\mathbf{H}} \in \mathbb{R}^{M\times M}, \tag{6}$$

$$\mathbf{X}_{inter}^{(l+1)} = \sigma(\mathbf{A}_{inter}\text{Mean}(\mathbf{X}_{intra}^{(l+1)})\mathbf{W}^{(l)}). \tag{7}$$

Specifically, the non-diagonal part of $\mathbf{A}'$ encode inter-subnetwork functional associations, i.e., $\mathbf{A}' \odot \mathbf{Mask}_{inter}$. A normalized binary incidence matrix $\hat{\mathbf{H}} = \mathbf{H}\mathbf{C}^{-1/2}$ mediates the scale transition, where $\mathbf{H} \in \{0, 1\}^{N\times M}$ satisfies $\mathbf{H}_{i,j} = 1 \Leftrightarrow v_i \in G_j$, and $\mathbf{C} = \text{diag}(\sum_i \mathbf{H}_{i,j})$ serves as the degree normalization matrix.

The result constitutes a compressed graph representation where each group-level node corresponds to a subnetwork. Group-level node features are computed through mean-pooling over constituent ROI features, followed by learnable transformations through a weight matrix. This hierarchical architecture enables simultaneous modeling of local functional units and global network interactions.

Inter-Intra Cross-Scale Message Passing. After obtaining the group-level feature representations $\mathbf{X}_{inter}^{(l+1)}$, the $\mathbf{X}_{inter}^{(l+1)}$ is passed back to the ROI-level features $\mathbf{X}_{intra}^{(l+1)}$ for cross-scale information fusion to realize the global feature update.

$$\mathbf{X}_{inter}^{(l+1)} \in \mathbb{R}^{M\times D} \longmapsto \Phi(\mathbf{X}_{inter}^{(l+1)}) \in \mathbb{R}^{N\times D}, \tag{8}$$

$$\mathbf{X}^{(l+1)} = \text{Fusion}(\mathbf{X}_{intra}^{(l+1)}, \mathbf{X}_{inter}^{(l+1)}). \tag{9}$$

Specifically, $\mathbf{X}_{inter}^{(l+1)}$ is first mapped by a mapping function $\Phi(\cdot)$ according to the brain region composition of each subnetwork, and then the ROI-level features $\mathbf{X}_{intra}^{(l+1)}$ and mapped group-level features $\Phi(\mathbf{X}_{inter}^{(l+1)})$ are added to perform feature fusion and global feature $\mathbf{X}^{(l+1)}$ update.

2.3 Readout Layer

The multi-scale features $\mathbf{X}^{(l+1)}$ are transferred to the Readout Layer. The readout result is generated by concat pooling operations and then fed into the MLP to obtain the result. The whole process is supervised with cross-entropy loss.

3 Experiments

3.1 Experimental Settings

Datasets and Preprocessing. The proposed DHGFormer is evaluated on two distinct brain disease diagnosis tasks. 1) *Autism Brain Imaging Data Exchange*(ABIDE)[20] aggregates neuroimaging data from 16 global research centers, comprising 539 individuals with ASD and 573 normal controls (NCs). 2) *Alzheimer's Disease Neuroimaging Initiative* (ADNI)[21] represents a multicenter longitudinal study focused on enhancing therapeutic intervention assessment for MCI. In this study, we select a subset of ADNI consisting of 125 MCI patients and 139 NCs. Data preprocessing is conducted using DPARSF[22].

Baselines. The DHGFormer is compared with several state-of-the-art methods, including GNN-based methods, Hierarchical-based methods, and transformer-based methods. The codes are reproduced based on the released codes.

Implement. The models are trained on an NVIDIA 3090 GPU. In the FC-Inspired encoder, the number of attention heads is set to 8. The training/test data is randomly split by five-fold cross-validation. During training, the Adam optimizer is utilized with an initial learning rate of 1e-4 and a weight decay factor of 1e-4. The batch size is 16 and the epoch is set to 100. The final results are expressed as mean values ± standard deviation of the five-fold cross-validation.

Evaluation Metrics. Within this research, all the experiments are structured as binary classification tasks. The classification performance is assessed using the following metrics: accuracy, sensitivity, specificity, and AUC.

3.2 Experimental Results Comparison and Analysis

As shown in Table 1, the DHGFormer achieves optimal performance across all metrics on the ABIDE and ADNI datasets. On the ABIDE dataset, the ACC of DHGFormer reaches 73.21%, which is 5.08% higher compared to the GNN-based method and 1.98% higher compared to the Transformer-based method; In the

Table 1. Experimental Results of the Comparison Methods.

Dataset	Method	ACC(%)	SEN(%)	SPE(%)	AUC
ABIDE	MVS-GCN[23]	$68.13_{\pm 2.64}$	$69.22_{\pm 5.21}$	$67.16_{\pm 4.73}$	$0.6804_{\pm 0.0286}$
	Com-BTF[16]	$70.60_{\pm 2.73}$	$71.99_{\pm 5.03}$	$69.38_{\pm 6.94}$	$0.7245_{\pm 0.0467}$
	FBNETGEN[11]	$70.46_{\pm 2.43}$	$69.79_{\pm 4.64}$	$71.16_{\pm 5.26}$	$0.7159_{\pm 0.0258}$
	HHGF[17]	$70.05_{\pm 3.47}$	$69.58_{\pm 5.47}$	$71.54_{\pm 6.26}$	$0.7126_{\pm 0.0547}$
	BrainIB[24]	$69.21_{\pm 8.64}$	$65.32_{\pm 5.70}$	$72.94_{\pm 5.32}$	$0.6902_{\pm 0.0327}$
	ALTER[25]	$71.23_{\pm 2.86}$	$71.27_{\pm 3.84}$	$71.22_{\pm 2.02}$	$0.7547_{\pm 0.0207}$
	DHGFormer	$\mathbf{73.21}_{\pm 1.51}$	$\mathbf{72.78}_{\pm 4.14}$	$\mathbf{74.02}_{\pm 3.80}$	$\mathbf{0.7646}_{\pm 0.0377}$
ADNI	MVS-GCN[23]	$63.67_{\pm 1.17}$	$65.30_{\pm 2.69}$	$61.78_{\pm 4.00}$	$0.6626_{\pm 0.0302}$
	Com-BTF[16]	$66.67_{\pm 3.72}$	$67.49_{\pm 2.67}$	$65.55_{\pm 3.56}$	$0.6880_{\pm 0.0301}$
	FBNETGEN[11]	$65.05_{\pm 1.86}$	$65.70_{\pm 2.50}$	$64.01_{\pm 5.83}$	$0.6690_{\pm 0.0332}$
	HHGF[17]	$67.35_{\pm 1.55}$	$66.28_{\pm 2.65}$	$68.44_{\pm 3.07}$	$0.7108_{\pm 0.0286}$
	BrainIB[24]	$64.68_{\pm 2.94}$	$66.36_{\pm 9.01}$	$63.62_{\pm 8.14}$	$0.6383_{\pm 0.0390}$
	ALTER[25]	$68.50_{\pm 1.17}$	$69.32_{\pm 2.28}$	$67.33_{\pm 2.58}$	$0.7200_{\pm 0.0181}$
	DHGFormer	$\mathbf{71.03}_{\pm 1.97}$	$\mathbf{71.83}_{\pm 2.25}$	$\mathbf{70.06}_{\pm 2.26}$	$\mathbf{0.7335}_{\pm 0.0352}$

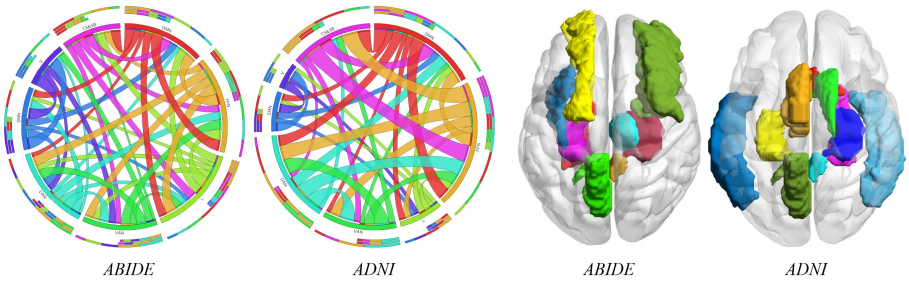
Table 2. Experimental Results of the Ablation Study.

Dataset	Method	ACC(%)	SEN(%)	SPE(%)	AUC
ABIDE	DHGFormer w/o dynamic	$67.88_{\pm 2.36}$	$68.39_{\pm 3.48}$	$67.50_{\pm 1.85}$	$0.7175_{\pm 0.0233}$
	DHGFormer w/o FC	$69.85_{\pm 1.72}$	$70.95_{\pm 3.52}$	$69.18_{\pm 3.30}$	$0.7311_{\pm 0.0140}$
	DHGFormer w/o HG	$69.65_{\pm 1.93}$	$68.21_{\pm 4.91}$	$71.24_{\pm 2.56}$	$0.7280_{\pm 0.0237}$
	DHGFormer	$\mathbf{73.21}_{\pm 1.51}$	$\mathbf{72.78}_{\pm 4.14}$	$\mathbf{74.02}_{\pm 3.80}$	$\mathbf{0.7646}_{\pm 0.0377}$
ADNI	DHGFormer w/o dynamic	$66.83_{\pm 1.79}$	$68.19_{\pm 2.63}$	$65.12_{\pm 2.23}$	$0.6894_{\pm 0.0215}$
	DHGFormer w/o FC	$69.49_{\pm 1.26}$	$69.41_{\pm 1.92}$	$69.58_{\pm 1.97}$	$0.7164_{\pm 0.0121}$
	DHGFormer w/o HG	$67.29_{\pm 2.54}$	$67.66_{\pm 1.83}$	$66.95_{\pm 1.67}$	$0.7008_{\pm 0.0397}$
	DHGFormer	$\mathbf{71.03}_{\pm 1.97}$	$\mathbf{71.83}_{\pm 2.25}$	$\mathbf{70.06}_{\pm 2.26}$	$\mathbf{0.7335}_{\pm 0.0352}$

ADNI dataset, the ACC (71.03%) and AUC (0.7335) of DHGFormer also outperform all baselines, with improvements of 2.53% and 1.35%, respectively, over the suboptimal method. Notably, DHGFormer maintains a balanced improvement in both SEN and SPE, demonstrating the effectiveness and robustness of DHGFormer. These notable gains validate our two core contributions: 1) Dynamic functional connectivity patterns guided by FC-inspired factors provide more discriminative network representations than static graphs; 2) Hierarchical modeling grounded in prior functional partition knowledge enables multi-scale feature integration from local functional subnetworks to the global brain.

3.3 Ablation Study

Three variants of the DHGFormer model are constructed to evaluate the role of the key module: DHGFormer w/o dynamic, replacing the dynamic generation method with the static PCC brain network construction method; DHGFormer w/o FC, removing the FC-inspired factors; and DHGFormer w/o HG, replacing the Hierarchical GNN instead of conventional GNN.



(a) Functional Connectivity at the Group-Level (b) Discriminative ROIs

Fig. 2. Visualization analysis of ROI-level and group-level associations.

As shown in Table 2, when the dynamic optimization mechanism of the brain network is removed, performance decreases by 5%. This marked decrease reveals that dynamically updating the brain network structure enables a more effective capture of the latent functional connectivity patterns associated with neurological disorders. Removal of the FC-inspired factor results in a performance reduction exceeding 3%. This finding suggests that FC as an attention guiding factor can effectively modulate the distribution of attention scores and prioritize biologically significant functional connections during brain network optimization. Moreover, replacing the hierarchical GNN with a conventional GNN causes over 3% performance degradation. This comparative analysis highlights the hierarchical design enables cross-scale feature integration from micro-level brain regions to macro-level functional subnetworks through prior knowledge incorporation, allowing comprehensive characterization of disease pathology.

3.4 Visualization

Our visual analysis reveals interaction patterns among functional subnetworks Fig. 2 (a), where connection width represents functional connectivity strength. Comparative analysis demonstrates that the Default Mode Network (DMN) exhibits significant hub properties in both neurodevelopmental disorders while displaying pathological heterogeneity in specific interaction patterns. Notably, the ABIDE dataset analysis shows pronounced DMN-Frontoparietal Network (FPN) connectivity, potentially associated with abnormal executive function regulation. Conversely, in the ADNI dataset, significant interaction occurs between DMN and the Dorsal Attention Network (DAN), suggesting compensatory alterations in sensorimotor integration mechanisms. Moreover, we further showcase the discriminative ROIs for disease diagnosis in Fig. 2 (b). These key ROIs mainly localize to DMN (ABIDE: Middle frontal gyrus, Posterior cingulate gyrus, ADNI: Hippocampus, Olfactory cortex) similar to the results of the group-level dependencies and consistent with the findings of previous studies [26,27,28].

4 Conclusion

This study presents a Dynamic Hierarchical Graph Transformer framework that advances brain network analysis by synergizing dynamic functional connectivity learning with hierarchical neural representation. By addressing the incompatibility of GNNs with PCC-based brain graphs, DHGFormer dynamically adapts edge weights to preserve neurobiologically critical inhibitory/excitatory interactions while integrating prior knowledge of functional subnetworks. Experimental results demonstrate the effectiveness of our method in both disease classification and interpretable biomarker discovery, outperforming compared methods. The framework's ability to reveal abnormal multi-scale connectivity patterns provides valuable insights into pathogenic mechanisms of neurological disorders, bridging the gap between data-driven deep learning and neuroanatomical priors.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant Nos. 62088102, 62125305, and 62401330, Guangdong Major Project of Basic and Applied Basic Research under Grant No. 2023B0303000009.

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

References

1. Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature reviews neuroscience*, 13(5):336–349, 2012.
2. Riitta Hari and Miiamaaria V Kujala. Brain basis of human social interaction: from concepts to brain imaging. *Physiological reviews*, 2009.
3. Yikai Wang and Ying Guo. A hierarchical independent component analysis model for longitudinal neuroimaging studies. *NeuroImage*, 189:380–400, 2019.
4. Roger E Beaty, Mathias Benedek, Paul J Silvia, and Daniel L Schacter. Creative cognition and brain network dynamics. *Trends in cognitive sciences*, 20(2):87–95, 2016.
5. Caio Seguin, Olaf Sporns, and Andrew Zalesky. Brain network communication: concepts, models and applications. *Nature reviews neuroscience*, 24(9):557–574, 2023.
6. Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. BrainGNN: Interpretable brain graph neural network for fMRI analysis. *Medical Image Analysis*, 74:102233, 2021.
7. Hejie Cui, Wei Dai, Yanqiao Zhu, Xuan Kan, Antonio Aodong Chen Gu, Joshua Lukemire, Liang Zhan, Lifang He, Ying Guo, and Carl Yang. Braingb: a benchmark for brain network analysis with graph neural networks. *IEEE transactions on medical imaging*, 42(2):493–506, 2022.
8. Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. Interpretable graph neural networks for connectome-based brain disorder analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 375–385. Springer, 2022.

9. Xiangmin Han, Rundong Xue, Shaoyi Du, and Yue Gao. Inter-intra high-order brain network for asd diagnosis via functional MRIs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 216–226. Springer, 2024.
10. Xiangmin Han, Rundong Xue, Jingxi Feng, Yifan Feng, Shaoyi Du, Jun Shi, and Yue Gao. Hypergraph foundation model for brain disease diagnosis. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.
11. Xuan Kan, Hejie Cui, Joshua Lukemire, Ying Guo, and Carl Yang. Fbnetgen: Task-aware gnn-based fmri analysis via functional brain network generation. In *International Conference on Medical Imaging with Deep Learning*, pages 618–637. PMLR, 2022.
12. Wenhao Jiang, Lin Ma, and Haifeng Li. Graph network modeling of brain connectivity: An exploration of word and object recognition tasks. In *2024 IEEE 17th International Conference on Signal Processing (ICSP)*, pages 692–696. IEEE, 2024.
13. Shengbo Tan, Rundong Xue, Shipeng Luo, Zeyu Zhang, Xinran Wang, Lei Zhang, Daji Ergu, Zhang Yi, Yang Zhao, and Ying Cai. Segkan: High-resolution medical image segmentation with long-distance dependencies. *arXiv preprint arXiv:2412.19990*, 2024.
14. Karl Friston. Hierarchical models in the brain. *PLoS computational biology*, 4(11):e1000211, 2008.
15. David Meunier, Renaud Lambiotte, Alex Fornito, Karen Ersche, and Edward T Bullmore. Hierarchical modularity in human brain functional networks. *Frontiers in neuroinformatics*, 3:571, 2009.
16. Anushree Bannadabhavi, Soojin Lee, Wenlong Deng, Rex Ying, and Xiaoxiao Li. Community-aware transformer for autism prediction in fMRI connectome. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 287–297. Springer, 2023.
17. Jiujiang Guo, Monan Wang, and Xiaojing Guo. Hierarchical hyperedge graph transformer: Towards dynamic interactions of brain networks for neurodevelopmental disease diagnosis. *IEEE Access*, 2024.
18. Ashish Vaswani et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, 2017.
19. BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 2011.
20. Cameron Craddock, Yassine Benhajali, Carlton Chu, Francois Chouinard, Alan Evans, András Jakab, Budhachandra Singh Khundrakpam, John David Lewis, Qingyang Li, Michael Milham, et al. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*, 7(27):5, 2013.
21. Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.
22. Chaogan Yan and Yufeng Zang. Dparsf: a matlab toolbox for" pipeline" data analysis of resting-state fmri. *Frontiers in systems neuroscience*, 4:1377, 2010.
23. Guangqi Wen, Peng Cao, Huiwen Bao, Wenju Yang, Tong Zheng, and Osmar Zaiane. Mvs-gcn: A prior brain structure learning-guided multi-view graph convo-

lution network for autism spectrum disorder diagnosis. *Computers in biology and medicine*, 142:105239, 2022.
24. Kaizhong Zheng, Shujian Yu, Baojuan Li, Robert Jenssen, and Badong Chen. Brainib: Interpretable brain network-based psychiatric diagnosis with graph information bottleneck. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.
25. Shuo Yu, Shan Jin, Ming Li, Tabinda Sarwar, and Feng Xia. Long-range brain graph transformer. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2025.
26. Hossein Shahamat and Mohammad Saniee Abadeh. Brain mri analysis using a deep learning based evolutionary approach. *Neural Networks*, 126:218–234, 2020.
27. Alexandre Abraham, Michael P Milham, Adriana Di Martino, R Cameron Craddock, Dimitris Samaras, Bertrand Thirion, and Gael Varoquaux. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage*, 147:736–745, 2017.
28. Andia H Turner, Kiefer S Greenspan, and Theo GM van Erp. Pallidum and lateral ventricle volume enlargement in autism spectrum disorder. *Psychiatry Research: Neuroimaging*, 252:40–45, 2016.