Adaptive Embedding for Long-Range High-Order Dependencies via Time-Varying Transformer on fMRI

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

[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

Abstract. Dynamic functional brain network analysis using rs-fMRI has emerged as a powerful approach to understanding brain disorders. However, current methods predominantly focus on pairwise brain region interactions, neglecting critical high-order dependencies and time-varying communication mechanisms. To address these limitations, we propose the Long-Range High-Order Dependency Transformer (LHDFormer), a neurophysiologically-inspired framework that integrates multiscale long-range dependencies with time-varying connectivity patterns. Specifically, we present a biased random walk sampling strategy with NeuroWalk kernel-guided transfer probabilities that dynamically simulate multi-step information loss through a k-walk neuroadaptive factor, modeling brain neurobiological principles such as distance-dependent information loss and state-dependent pathway modulation. This enables the adaptive capture of the multi-scale short-range couplings and long-range high-order dependencies corresponding to different steps across evolving connectivity patterns. Complementing this, the time-varying transformer co-embeds local spatial configurations via topology-aware attention and global temporal dynamics through cross-window token guidance, overcoming the single-domain bias of conventional graph/transformer methods. Extensive experiments on ABIDE and ADNI datasets demonstrate that LHDFormer outperforms state-of-the-art methods in brain disease diagnosis. Crucially, the model identifies interpretable high-order connectivity signatures, revealing disrupted long-range integration patterns in patients that align with known neuropathological mechanisms.

Keywords: Brain Network · Long-Range Dependencies · Brain Disease.

1 Introduction

Dynamic functional brain network analysis based on rs-fMRI data serves as a cornerstone for decoding human brain organization and identifying pathological sig-

natures in neurological disorders [1–3]. Mounting evidence highlights that time-varying functional connectivity (FC) encodes critical information about evolving neural coordination mechanisms underlying cognitive processes and neuropsychiatric disorders. However, current dynamic functional brain network analysis paradigms mainly focus on pairwise region-of-interest (ROI) interactions while neglecting the long-range high-order dependencies [4–7]. This oversight fundamentally limits the ability to model the long-range communication mechanisms within the brain, where high-order functional integration across distributed regions dynamically coordinates complex cognitive functions [8].

Existing methods for modeling dynamic brain networks typically analyze pairwise relationships through sliding window correlations or time-frequency coherence, failing to capture high-order information flow beyond immediate neighbors. Though recent studies [9,10] introduce random-walk kernels to model long-range dependencies, the designed static kernels exhibit two critical shortcomings: 1) inability to adapt to the time-varying nature of functional connectivity and 2) failure to encode neurophysiological multi-scale long-range dependency patterns corresponding to different steps. These limitations underscore an urgent need for frameworks that simultaneously address the *time-varying adaptivity* and *high-order dependency* challenges in dynamic brain network analysis.

To address the above challenges, we propose LHDFormer–a neurophysiologically grounded framework that combines neuroadaptive long-range dependency embedding with temporal dynamic integration. Specifically, we develop a biased random walk sampling strategy with a time-varying NeuroWalk kernel that enables dynamically regulating the precess of multi-step information propagation to generate long-range high-order dependency embeddings. Subsequently, local spatial dependencies within the brain and global dynamic connectivity patterns are integrated via a time-varying transformer based on the long-range embeddings. The main contributions of this work are as follows:

1) We propose a novel neuroadaptive long-range dependency awareness mechanism, incorporating a NeuroWalk kernel with biased random walks to guide transfer probabilities to capture multi-scale dependency patterns. This mechanism enables the simultaneous encoding of short-range functional couplings and long-range high-order dependencies within brain networks.
2) We develop a Time-Varying Transformer framework that integrates the local spatial topologies with evolving temporal dependencies through: Topology-aware attention preserving spatial neighborhood relationships, and Cross-window token integration capturing global connectivity evolution.
3) Extensive experiments on ABIDE and ADNI datasets demonstrate that LHDFormer outperforms all comparison methods. Visualization of the high-order dependencies reveals the brain's long-range discriminative patterns for brain diseases, offering new perspectives on their pathological mechanisms.

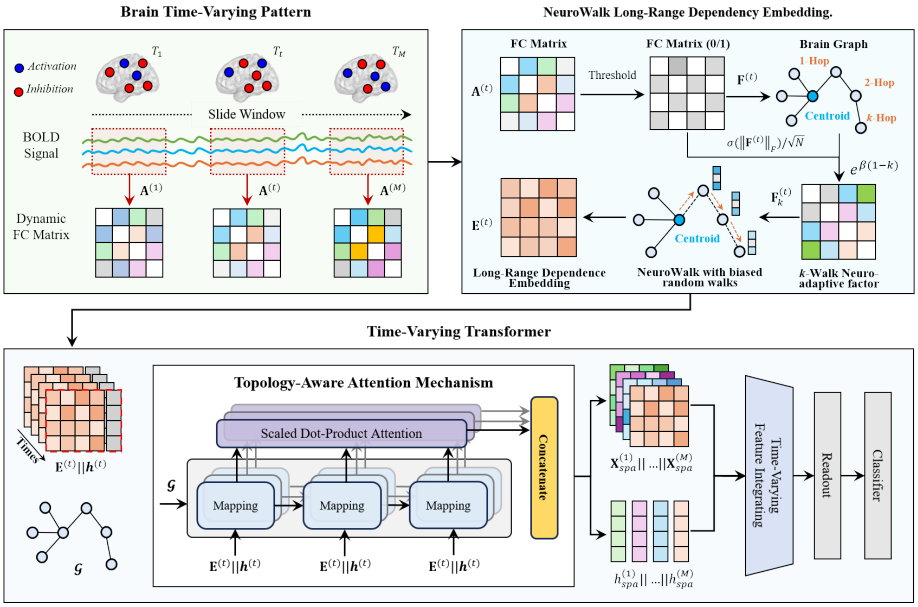


Fig. 1. The framework of the proposed LHDFormer.

2 Method

2.1 Adaptive Embedding for Long-Range High-Order Dependencies

Existing approaches primarily aggregate information from pairwise ROIs, neglecting the long-range high-order dependencies among ROIs. Although Study [9] introduced a long-range random walk kernel to address this limitation, its static design proves inadequate for temporal data like fMRI. Specifically, the static random walk kernel fails to discern distinct regulatory requirements across varying walk lengths and cannot capture temporal dynamic patterns, constraining the representational capacity of long-range embeddings. We develop a neuroadaptive long-range dependency embedding method that incorporates a NeuroWalk kernel with biased random walks to guide transfer probabilities. Through allocating independent adaptive factors for each walking step, our method achieves differentiated modulation of dependencies at varying distances, enabling the discovery of long-range high-order functional dependencies within the brain network.

Time-Varying Factors. For fMRI data, the BOLD time series are extracted across N ROIs for T time points. The temporal variability of functional connectivity between ROIs is represented using the sliding window technique, which segments the time series into M non-overlapping windows. The dynamic brain network is formalized as a sequence $\mathcal{G}^{(t)} = \{\mathcal{V}, \mathbf{A}^{(t)}\}_{t=1}^{M}$, where $\mathbf{A}^{(t)}$ is derived from the PCC matrix within a temporal window of length T/M. These time-varying adjacency matrices are thresholded as time-varying factors $\mathbf{F}^{(t)}$, with

each factor encoding inter-regional interactions during specific temporal phases, enabling longitudinal monitoring of evolving neural coordination patterns [11].

***k*-Walk Neuroadaptive Factors.** To capture multi-scale/multi-step long-range dependency patterns and simulate information decay during traversals, we propose a k-walk adaptive factor allocation mechanism that independently modulates information propagation intensity at different walk steps. For each random walk step $k \in \{1, \ldots, K\}$ at time t, step-specific adaptive factors $\mathbf{F}_k^{(t)}$ are generated through learnable transformation of functional connectivity patterns:

$$\mathbf{F}_k^{(t)} = e^{\beta(1-k)} \cdot \sigma(\| \mathbf{F}^{(t)} \|_F / \sqrt{N}), \tag{1}$$

where $\beta \in (0, 1)$ governs the base decay rate modeling biological signal loss, while the Frobenius norm $\| \mathbf{F}^{(t)} \|_F$ quantifies global connection intensity within the window t. This dual-constraint design ensures: 1) Exponential decay $e^{\beta(1-k)}$ simulates neurophysiological information loss; 2) Time-aware scaling adaptively modulates decay rates according to time-varying brain network density. The resultant factors automatically suppress noisy propagation in low-synchrony phases $|\mathbf{F}^{(t)}|_F \to 0$ while preserving salient patterns in high-coherence states, achieving biologically plausible dynamic filtering across spatiotemporal domains.

This k-walk neuroadaptive factor design enables differentiated regulation of biologically multi-scale dependencies: lower k-walks prioritize direct interactions, while higher k-walks selectively retain salient long-range patterns.

NeuroWalk Long-Range Dependency Embedding. The pre-computed neuroadaptive factors $\mathbf{F}_k^{(t)}$ effectively represent the multi-level long-range dependency patterns among ROIs corresponding to multi-steps. To encode these multi-scale relationships with time-varying properties, we implement a biased random walk sampling strategy with NeuroWalk kernel-guided transfer probabilities.

In conventional network theory, random walks follow Markovian transitions where the probability of moving from node A to B depends solely on the immediate network state, as defined by the transfer probability matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$. However, functional brain networks exhibit distinctive neurophysiological properties characterized by dynamic communication strengths between ROI pairs. Improper handling of these time-varying interactions may disrupt the intrinsic collaborative dynamics in brain cognitive processes. Given this fact, the transfer probability matrix within the brain network can be fine-tuned by k-walk neuroadaptive factors $\mathbf{F}_k^{(t)}$: $\hat{\mathbf{P}}_k^{(t)} = \mathbf{P} \odot \mathbf{F}_k^{(t)}$. Formally, given the adjacency matrix $\mathbf{A}^{(t)}$ and its degree diagonal matrix $(\mathbf{D}^{(t)})^{-1}$ at time t, along with the k-walk adaptive factor $\mathbf{F}_k^{(t)}$, we define the time-varying NeuroWalk kernel $\mathbf{R}^{(t,k)}$ as:

$$\mathbf{R}^{(t,k)} = (\mathbf{F}_k^{(t)} \odot \mathbf{A}^{(t)})(\mathbf{D}^{(t)})^{-1}. \tag{2}$$

This kernel enables progressive exploration of high-order dependencies through multi-step walks. For k-step random walks, the long-range embeddings $\mathbf{E}^{(t)}$ is

constructed through iterative aggregation:

$$\boldsymbol{e}_i^{(t)} = \left[\mathbf{I}, \mathbf{R}^{(t,1)}, \prod_{k=1}^{2} \mathbf{R}^{(t,k)}, \ldots, \prod_{k=1}^{K-1} \mathbf{R}^{(t,k)}\right]_{i,i}. \tag{3}$$

2.2 Time-Varying Transformer

To comprehensively model evolving neurocognitive patterns, we propose a Time-Varying Transformer that synergistically integrates local spatial dependencies and global dynamic connectivity patterns through dual complementary encoding pathways. The composite embedding at time window t is formulated as: $\mathbf{X}^{(t)} = (\mathbf{E}^{(t)})$. This framework establishes global feature representations through dual-domain processing that simultaneously captures intra-window spatial relationships and inter-window temporal dynamics.

Local Spatial-Domain Feature Encoding. The Time-Varying Transformer layer enhances local feature representation through a topology-aware attention mechanism that jointly processes local composite embeddings the local composite embeddings $\mathbf{X}^{(t)}$ and learnable spatial tokens $\boldsymbol{h}^{(t)}$ at time step t:

$$(\mathbf{X}^{(t)}, \boldsymbol{h}^{(t)}) \rightarrow (\mathbf{X}_{spa}^{(t)}.\boldsymbol{h}_{spa}^{(t)}). \tag{4}$$

The topology-aware attention mechanism preserves spatial graph properties and semantic attention patterns through the integration of graph convolution operators into attention computation, allowing the obtained attention scores to represent both functional node similarity and graph connectivity.

$$Q_i^{(t)}, K_i^{(t)}, V_i^{(t)} = \mathrm{ReLU}((\mathbf{D}^{(t)})^{-1/2}\mathbf{A}^{(t)}(\mathbf{D}^{(t)})^{-1/2}\mathbf{X}^{(t)})(W_Q, W_K, W_V), \tag{5}$$

$$\mathbf{X}_{spa}^{(t)} = W_o(\|_{i=1}^{D} \mathbf{X}_i^{(t)}), \quad \mathbf{X}_i^{(t)} = \mathrm{softmax}\Big(\frac{Q_i^{(t)}(K_i^{(t)})^\top}{\sqrt{d}}\Big)V_i^{(t)}, \tag{6}$$

where D is the number of attention heads, $\|$ indicates concatenation.

Global Time-Varying Feature Integrating. Following spatial encoding through l stacked layers, the temporal evolution of composite features is captured in the token sequence $\{\boldsymbol{h}_{spa}^{(t)}\}_{t=1}^{M}$. These encoded features across time $\|_{t=1}^{M} \mathbf{X}_{spa}^{(t)}$ are globally in the temporal domain under the guidance of token sequence $\|_{t=1}^{M} \boldsymbol{h}_{spa}^{(t)}$ of a standard Transformer, culminating in a final $\mathbf{X}_{tem}^{(l+1)}$ used for classification.

$$(\|_{t=1}^{M} \mathbf{X}_{spa}^{(t)}, \|_{t=1}^{M} \boldsymbol{h}_{spa}^{(t)}) \rightarrow \mathbf{X}_{tem}. \tag{7}$$

2.3 Readout Layer

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

Table 1. Experimental Results of the Comparison Methods.

Dataset	Method	ACC(%)	SEN(%)	SPE(%)	AUC
ABIDE	STGCN [15]	$65.52_{\pm 2.46}$	$62.96_{\pm 7.51}$	$67.74_{\pm 5.00}$	$0.6628_{\pm 0.0243}$
	BrainIB [16]	$69.21_{\pm 8.64}$	$65.32_{\pm 5.70}$	$72.94_{\pm 5.32}$	$0.6902_{\pm 0.0327}$
	RGTNET [17]	$69.75_{\pm 1.41}$	$70.30_{\pm 4.71}$	$68.87_{\pm 3.84}$	$0.7051_{\pm 0.0117}$
	MSSTAN [7]	$71.40_{\pm 1.49}$	$70.35_{\pm 1.80}$	$72.55_{\pm 1.83}$	$0.7678_{\pm 0.0076}$
	ALTER [9]	$71.23_{\pm 2.86}$	$71.27_{\pm 3.84}$	$71.22_{\pm 2.02}$	$0.7547_{\pm 0.0207}$
	LHDFormer	$\mathbf{74.29}_{\pm 1.17}$	$\mathbf{73.48}_{\pm 1.10}$	$\mathbf{75.09}_{\pm 2.04}$	$\mathbf{0.7978}_{\pm 0.0103}$
ADNI	STGCN [15]	$62.82_{\pm 4.46}$	$65.23_{\pm 5.49}$	$60.36_{\pm 4.65}$	$0.6254_{\pm 0.0582}$
	BrainIB [16]	$64.68_{\pm 2.94}$	$66.36_{\pm 9.01}$	$63.62_{\pm 8.14}$	$0.6583_{\pm 0.0390}$
	RGTNET [17]	$65.75_{\pm 3.23}$	$66.32_{\pm 4.28}$	$64.54_{\pm 3.15}$	$0.6826_{\pm 0.0449}$
	MSSTAN [7]	$67.88_{\pm 2.84}$	$69.38_{\pm 3.83}$	$66.40_{\pm 2.73}$	$0.7130_{\pm 0.0457}$
	ALTER [9]	$68.50_{\pm 1.17}$	$69.32_{\pm 2.28}$	$67.33_{\pm 2.58}$	$0.7200_{\pm 0.0181}$
	LHDFormer	$\mathbf{71.42}_{\pm 1.70}$	$\mathbf{72.52}_{\pm 2.02}$	$\mathbf{70.27}_{\pm 1.10}$	$\mathbf{0.7347}_{\pm 0.0209}$

3 Experiments

3.1 Experimental Settings

Datasets and Preprocessing. We evaluate LHDFormer on two distinct brain disease diagnosis tasks: 1) *Autism Brain Imaging Data Exchange*(ABIDE) [12] aggregates neuroimaging data across 16 international sites, comprising 539 individuals with ASD and 573 normal controls (NCs). 2) *Alzheimer's Disease Neuroimaging Initiative* (ADNI) [13] 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. All fMRI data underwent standardized preprocessing through DPARSF [14].

Compared Methods. The LHDFormer is compared with several state-of-the-art methods, including GNN-based methods and transformer-based methods. The codes are reproduced based on the released codes or detailed introductions.

Implement. *Device*: Experiments are conducted on an NVIDIA RTX 3090 GPU using PyTorch. *Data split*: Stratified 5-fold cross-validation (4:1 train/test split). *Optimization*: Adam with learning rate 1e-4, weight decay 1e-4. *Training*: 100 epochs with 16 batch size, early stopping. The final results are expressed as mean values ± standard deviation of the five-fold cross-validation.

Evaluation Metrics. For binary neurodiagnostic classification tasks, the classification performance is assessed using the following metrics: accuracy, sensitivity, specificity, and Area Under the Curve (AUC).

3.2 Experimental Results and Discussion

As detailed in Table 1, LHDFormer establishes state-of-the-art performance on both benchmark datasets. On ABIDE, our method achieves 74.29% accuracy,

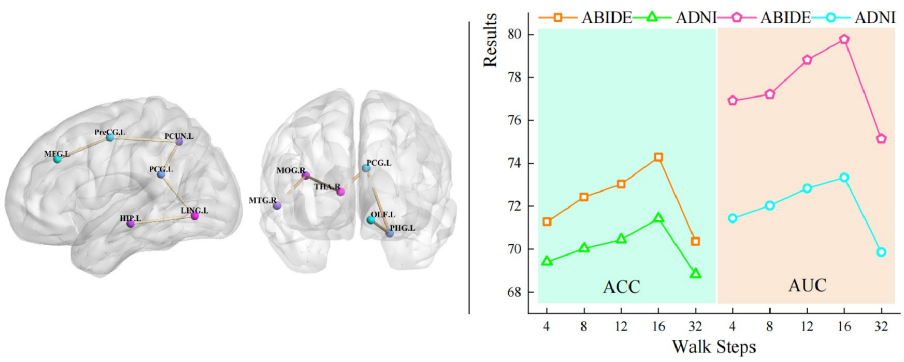
Table 2. Experimental Results of the Ablation Study.

Dataset	Method	ACC(%)	SEN(%)	SPE(%)	AUC
ABIDE	LHDFormer w/o Emb	$69.56_{\pm 1.03}$	$69.34_{\pm 1.61}$	$69.75_{\pm 2.38}$	$0.7509_{\pm 0.0103}$
	LHDFormer w/o NeuroWK	$71.64_{\pm 0.91}$	$71.96_{\pm 1.13}$	$71.33_{\pm 1.88}$	$0.7668_{\pm 0.0189}$
	LHDFormer w/o TopoAM	$72.12_{\pm 1.41}$	$71.32_{\pm 1.43}$	$72.81_{\pm 2.96}$	$0.7726_{\pm 0.0112}$
	LHDFormer	$\mathbf{74.29}_{\pm 1.17}$	$\mathbf{73.48}_{\pm 1.10}$	$\mathbf{75.09}_{\pm 2.04}$	$\mathbf{0.7978}_{\pm 0.0103}$
ADNI	LHDFormer w/o Emb	$65.20_{\pm 1.17}$	$67.63_{\pm 2.98}$	$64.82_{\pm 2.43}$	$0.6986_{\pm 0.0232}$
	LHDFormer w/o NeuroWK	$67.58_{\pm 2.34}$	$68.76_{\pm 2.05}$	$66.56_{\pm 2.92}$	$0.7084_{\pm 0.0159}$
	LHDFormer w/o TopoAM	$69.19_{\pm 1.34}$	$69.71_{\pm 1.96}$	$68.85_{\pm 2.50}$	$0.7189_{\pm 0.0196}$
	LHDFormer	$\mathbf{71.42}_{\pm 1.70}$	$\mathbf{72.52}_{\pm 2.02}$	$\mathbf{70.27}_{\pm 1.10}$	$\mathbf{0.7347}_{\pm 0.0209}$

outperforming GNN baselines and transformer competitors. Particularly noteworthy are the consistent gains in SEN and SPE, coupled with a 0.0431 AUC improvement compared to the other baselines. For ADNI, LHDFormer attains 71.42% accuracy and 0.7347 AUC, surpassing the suboptimal method by 2.92% and 0.0147 respectively, while maintaining balanced SEN/SPE enhancements. This superiority stems from: 1) the Time-varying NeuroWalk kernel simulates multi-step information transfer mechanisms in the brain to capture long-range high-order dependencies within brain networks; 2) the Time-Varying Transformer's synergistic encoding of local spatial patterns (through topology-aware attention mechanism) and global temporal dynamics (via cross-window token guidance) overcomes the single-domain bias in GNNs/Transformers.

3.3 Ablation Study

To evaluate the contribution of core components, we develop three variants of the LHDFormer model: LHDFormer w/o Emb, which eliminates the long-range embeddings; LHDFormer w/o NeuroWK, substituting our time-varying NeuroWalk kernel with a conventional random-walk kernel [9]; and LHDFormer w/o TopoAM, replacing the topology-aware attention with standard self-attention. 1) As shown in Table 2, removing the long-range embeddings results in accuracy reductions of 4.73% and 6.22% on the ABIDE and ADNI datasets, respectively. This substantial performance degradation underscores the critical role of long-range embeddings in capturing global high-order dependencies. 2) Replacing the time-varying NeuroWalk kernel with a static random-walk kernel leads to accuracy declines of 2.65% (ABIDE) and 3.84% (ADNI). This validates the NeuroWalk kernel's adaptive modeling of the temporal evolution of long-range correlations, enabling the model to capture disease-related alterations in latent connectivity patterns. 3) Moreover, ablating the topology-aware attention module causes performance deterioration, demonstrating that the topology-aware attention mechanism enables simultaneous learning of node-level interactions and preservation of graph structural information compared with self-attention.



(a) Long-Range High-Order Dependencies (b) Influence of Walk Steps

Fig. 2. Further Discussion of Long-Range Higher-Order Dependencies

3.4 Further Discussion

We first visualize discriminative long-range dependencies across both datasets (shown as Figure. 2(a)). For the ABIDE dataset, the identified long-range dependencies outlined ASD pathology pathways, including prefrontal executive control, sensorimotor integration, and visual memory compensation mechanisms. Observed connectivity is associated with ASD: Diminished functional coupling between MFG and PCUN correlates with social impairment, dysregulated connectivity between PCG and LING corresponds to abnormal sensory processing, and altered interactions between PreCG and HIP may underlie stereotypic behavior [18–20]. Regarding the ADNI dataset, the OLF-PHG-THA-PCG-MOG-MTG connectivity pattern reflects the multifaceted pathological processes of MCI: Early MCI pathology spreads through OLF-PHG, consistent with Braak staging of AD; Visual semantic network dysfunction (MOG-MTG) leads to complex task deficits by disrupting visual-memory interactions [21–23].

To optimize model performance, we investigate the Walk Step hyperparameter (steps=4, 8, 12, 16, 32) on both datasets. As shown in Figure. 2(b), predictive accuracy exhibits a roughly positive correlation with increasing Walk Step values. This phenomenon may reflect the neurobiological reality of distributed information transfer across distal brain regions. LHDFormer effectively captures these long-range dependencies, enhancing its diagnostic utility for brain disorders.

4 Conclusion

This study presents LHDFormer, a neurophysiologically inspired framework for dynamic high-order brain network analysis. By integrating a time-varying NeuroWalk kernel with a dual-domain time-varying transformer, our method achieves adaptive modeling of long-range high-order dependencies while preserving the dynamic nature of functional connectivity. The NeuroWalk kernel, driven by k-walk neuroadaptive adaptive factors, enables adaptively capturing different step

multi-scale dependency patterns through signal decay and synchronization-aware modulation. Coupled with the Time-Varying Transformer, the framework jointly encodes local spatial features and global temporal coordination patterns, establishing a unified representation of brain network dynamics. Future directions include extending LHDFormer to multi-center cohorts to enhance generalizability and refining the adaptive factor mechanism using biophysical constraints.

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. Partha P Mitra and Bijan Pesaran. Analysis of dynamic brain imaging data. *Biophysical journal*, 76(2):691–708, 1999.
2. Karl J Friston, Christopher D Frith, Richard SJ Frackowiak, and Robert Turner. Characterizing dynamic brain responses with fmri: a multivariate approach. *Neuroimage*, 2(2):166–172, 1995.
3. Mikhail I Rabinovich and Pablo Varona. Robust transient dynamics and brain functions. *Frontiers in computational neuroscience*, 5:24, 2011.
4. Lanting Li, Liuzeng Zhang, Peng Cao, Jinzhu Yang, Fei Wang, and Osmar R Zaiane. Exploring spatio-temporal interpretable dynamic brain function with transformer for brain disorder diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 195–205. Springer, 2024.
5. 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.
6. Byung-Hoon Kim, Jungwon Choi, EungGu Yun, Kyungsang Kim, Xiang Li, and Juho Lee. Learning dynamic brain connectome with graph transformers for psychiatric diagnosis classification. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2024.
7. Youyong Kong, Xiaotong Zhang, Wenhan Wang, Yue Zhou, Yueying Li, and Yonggui Yuan. Multi-scale spatial-temporal attention networks for functional connectome classification. *IEEE Transactions on Medical Imaging*, 2024.
8. 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.
9. 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.
10. 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.

11. Biao Jie, Mingxia Liu, and Dinggang Shen. Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease. *Medical image analysis*, 47:81–94, 2018.
12. 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.
13. 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.
14. Chaogan Yan and Yufeng Zang. Dparsf: a matlab toolbox for" pipeline" data analysis of resting-state fmri. *Frontiers in systems neuroscience*, 4:1377, 2010.
15. Soham Gadgil, Qingyu Zhao, Adolf Pfefferbaum, Edith V Sullivan, Ehsan Adeli, and Kilian M Pohl. Spatio-temporal graph convolution for resting-state fmri analysis. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23*, pages 528–538. Springer, 2020.
16. 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.
17. Yibin Wang, Haixia Long, Tao Bo, and Jianwei Zheng. Residual graph transformer for autism spectrum disorder prediction. *Computer Methods and Programs in Biomedicine*, 247:108065, 2024.
18. Catherine E Rice, Michael Rosanoff, Geraldine Dawson, Maureen S Durkin, Lisa A Croen, Alison Singer, and Marshalyn Yeargin-Allsopp. Evaluating changes in the prevalence of the autism spectrum disorders (asds). *Public health reviews*, 34:1–22, 2012.
19. Ann Genovese and Merlin G Butler. Clinical assessment, genetics, and treatment approaches in autism spectrum disorder (asd). *International journal of molecular sciences*, 21(13):4726, 2020.
20. Catherine Lord, Mayada Elsabbagh, Gillian Baird, and Jeremy Veenstra-Vanderweele. Autism spectrum disorder. *The lancet*, 392(10146):508–520, 2018.
21. Florence Portet, PJ Ousset, PJ Visser, GB Frisoni, F Nobili, Ph Scheltens, B Vellas, J Touchon, MCI Working Group of the European Consortium on Alzheimer's Disease (EADC, et al. Mild cognitive impairment (mci) in medical practice: a critical review of the concept and new diagnostic procedure. report of the mci working group of the european consortium on alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 77(6):714–718, 2006.
22. Howard H Feldman and Claudia Jacova. Mild cognitive impairment. *The American Journal of Geriatric Psychiatry*, 13(8):645–655, 2005.
23. Ronald C Petersen. Mild cognitive impairment. *CONTINUUM: lifelong Learning in Neurology*, 22(2):404–418, 2016.