# Heterogeneous Masked Attention-Guided Path Convolution for Functional Brain Network Analysis

Jiakun Xu[1], Xin Zhang[1(✉)], Tong Xiong[1], Shengxian Chen[1], Xiaofen Xing[1], Jindou Hao[2], and Xiangmin Xu[1(✉)]

[1] School of Electronic and Information Engineering, South China University of Technology, Guangzhou, Guangdong, China
[2] Department of Pediatrics, Shenzhen Maternity and Child Healthcare Hospital, Southern Medical University, Shenzhen, Guangdong, China
{eexinzhang,xmxu}@scut.edu.cn

**Abstract.** Brain functional connectivity analysis plays a crucial role in the computer-aided diagnosis of brain disorders. The brain organization is a heterogeneous structure with distinct functional divisions. However, current heterogeneous algorithms often introduce excessive parameters while characterizing heterogeneous relationships, leading to redundancy and overfitting. To address these issues, we propose the Heterogeneous Masked Attention-Guided Path Convolution (HM-AGPC) for functional brain network analysis. The HM-AGPC introduces a heterogeneous masked attention generation mechanism that preserves valuable heterogeneous relationships while minimizing redundant interactions and highlighting crucial functional connections. Moreover, the framework incorporates an attention-guided path convolution strategy, which leverages attention weights to guide the convolution kernel in focusing on the most salient features and pathways. This approach improves model performance without directly introducing extra parameters, thereby enhancing feature learning efficiency. We evaluate HM-AGPC on the ABIDE dataset using ten-fold cross-validation, where it demonstrates superior performance in disease diagnosis task compared to state-of-the-art methods. Additionally, the framework demonstrates high interpretability, making it a promising tool for computer-aided diagnosis and the identification of potential biomarkers.

**Keywords:** Heterogeneous mask · Attention-guided path convolution · Functional brain network analysis.

## 1 Introduction

The resting-state functional Magnetic Resonance Imaging (rs-fMRI) measures blood oxygen level-dependent (BOLD) signals by capturing the content alteration of oxyhemoglobin and deoxyhemoglobin. Reflecting the interactions between neuron clusters, rs-fMRI signals provide insights into brain's activation

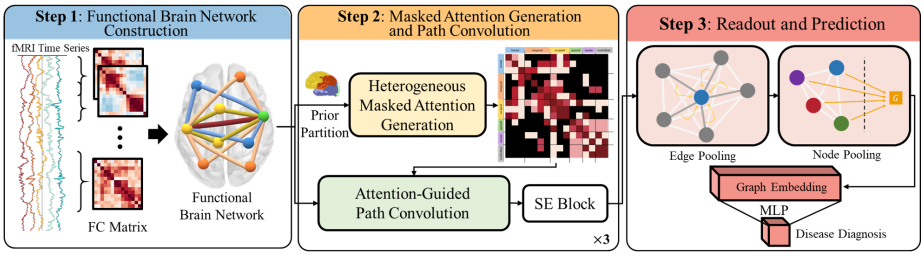


**Fig. 1.** The overview of HM-AGPC. Step 1 (blue): Functional Brain Network Construction. Step 2 (orange): Masked Attention Generation and Path Convolution, including our two key components: Heterogeneous Masked Attention Generation and Attention-Guided Path Convolution. Step 3 (red): Readout and Prediction.

patterns [1]. It has been widely used for brain researches because of its advantage of non-invasive and high-resolution [2–4]. Cognitive function arises from intricate interactions among clusters of densely interconnected brain regions, which form the complex community architecture of human brain networks [5].

The rs-fMRI based brain network can be mathematically considered as a fully connected graph, where brain regions are nodes and functional interactions are edges. With the development of deep learning, many graph models show significant feature extraction capabilities in brain networks, especially graph neural networks (GNN) [6, 7]. However, the region-wise interaction information of the brain network is contained in the edges, which limits the direct application of the node-based GNN. Then, several edge graph models have been proposed [8–10]. By using different convolution mechanisms to aggregate edge features, they have shown improved performance in age estimation and disease diagnosis.

These algorithms all treat brain functional networks as homogeneous graphs, whereas many studies have demonstrated that brain organization is a heterogeneous structure with distinct functional divisions [11,12]. Consequently, graph algorithms based on heterogeneous brain networks have emerged [13,14]. Through the use of heterogeneous networks, these methods can characterize diverse brain connectivity relationships more accurately, achieving promising results in brain disease analysis and brain development studies. However, these algorithms introduce an excessive number of parameters during the process of heterogeneous convolution, which can easily lead to severe overfitting in brain network analysis.

Transformer-based methods are increasingly being applied to the analysis of functional brain networks [15–17]. Leveraging the attention mechanism [18], these methods excel at capturing global information, thereby achieving impressive performance. However, these approaches often lack detailed masking of the attention matrices and overlook the heterogeneity of brain connectivity, which limits their ability to minimize less significant interactions and to accurately model functional brain networks.

To address these issues, we propose a Heterogeneous Masked Attention-Guided Path Convolution (HM-AGPC) for functional brain network analysis.

The proposed HM-AGPC incorporates a heterogeneous masked attention generation mechanism that can preserve valuable heterogeneous relationships and emphasize crucial connections. Additionally, it includes an attention-guided path convolution that enables the convolution kernels to capture salient features effectively without directly introducing any additional parameters. The major contributions of this work are highlighted below:

1 We propose a novel heterogeneous masking mechanism to characterize the heterogeneity of functional brain networks. This mechanism selectively emphasizes valuable heterogeneous relationships while minimizing redundant interactions, thereby enhancing the model's ability to highlight crucial connections.

2 We adopt a concise attention-guided path convolution strategy. Without directly introducing additional parameters, it can guide the graph path convolution to learn more effective convolution kernels, thereby enhancing the network's ability to capture key pathways and dynamic relationships among brain regions.

3 We conduct experiments on the ABIDE dataset to demonstrate the superiority of our method over other state-of-the-art approaches. Furthermore, our method exhibits strong interpretability, offering a novel approach for computer-aided diagnosis of brain disorders and the identification of potential biomarkers.

## 2 Method

The proposed HM-AGPC framework is illustrated in Fig.1. Below, we first introduce the step 1: Functional Brain Network Construction. Then, we introduce two key modules in the step 2: Heterogeneous Masked Attention Generation and Attention-Guided Path Convolution. Finally, we provide a concise description of the step 3: Readout and Prediction.

### 2.1 Functional Brain Network Construction

To construct brain networks, a brain atlas is utilized to define a set of nodes, each representing a Region of Interest (ROI). The rs-fMRI BOLD time series for each ROI is computed by averaging the time series of all voxels within the ROI. For each subject, we obtain $N$ time series corresponding to $N$ ROIs. A sliding window approach is applied to segment time series into multiple overlapping slices with a fixed stride. For each slice, pairwise Pearson's correlation coefficients are calculated between ROIs to construct a functional connectivity (FC) matrix.

The brain network can be represented as a fully connected graph, denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_i\}_{i=1}^{N} \in \mathbb{R}^N$ represents the set of nodes (ROIs), and $\mathcal{E} = \{e_{ij}\} \in \mathbb{R}^{N \times N}$ represents edges which are the connections of nodes defined by FC. Edges are initialized with D-dimensional edge features $\mathbf{E} \in \mathbb{R}^{N \times N \times D}$.

### 2.2 Heterogeneous Masked Attention Generation

The interaction patterns among brain regions exhibit inherent differences. To effectively capture these heterogeneous relations, we introduce a region-based heterogeneous mask strategy. The process is shown in the yellow part of Fig.2.

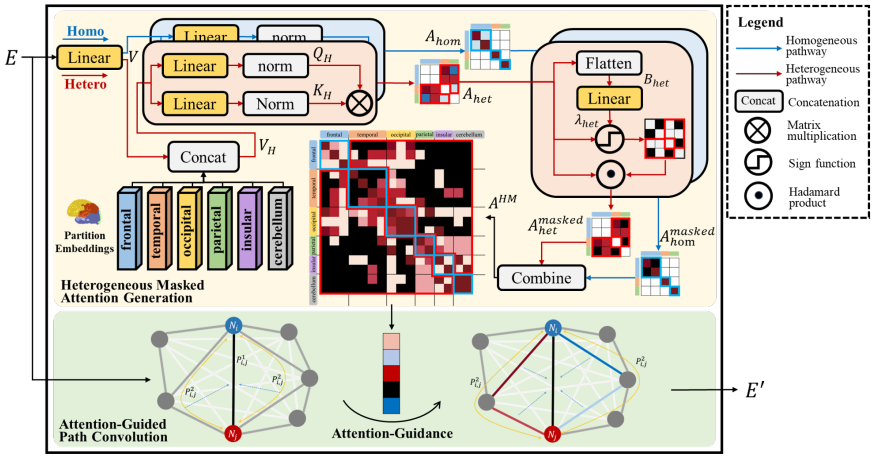


**Fig. 2.** The diagram of Heterogeneous Masked Attention Generation (yellow) and Attention-Guided Path Convolution (green). The heterogeneous and homogeneous pathways are marked with red and blue respectively.

**Partition Embedding Concatenations.** First, we map the edge features $\mathbf{E}$ to node features $\mathbf{V}$ through a linear layer. Next, we adopt a pre-defined brain partition to categorize $N$ brain regions into $T$ distinct groups. We generate a special partition embedding $H_t$ for each category and concatenate the node feature $V_i$ with its corresponding partition embeddings to obtain the extended node features $V_H$:

$$\mathbf{V} = \mathbf{E}w_N, V_{i_H} = \text{Concat}(V_i, H_t), \tag{1}$$

where $w_N \in \mathbb{R}^{N\times 1}$ is the learnable parameter, $\mathbf{V} \in \mathbb{R}^{N\times D}$ is the node feature, $H_t \in \mathbb{R}^{D_H\times 1}$ is the partition embedding of the $t^{th}$ group. $V_i \in \mathbb{R}^{D\times 1}$ is the feature vector of node $v_i$ which is assigned to the same $t^{th}$ group according to the previous partition scheme, i.e. $Type(v_i) = t$, and $V_{i_H} \in \mathbb{R}^{(D+D_H)\times 1}$ is the extended feature of $v_i$.

**Hetero- and Homo- Attention Matrix Calculation.** Given that the interactions between heterogeneous nodes and homogeneous nodes are distinct, from this step onward, we process heterogeneous and homogeneous parts separately. Specifically, we calculate the hetero- and homo- attention matrix ($\mathbf{A}_{\text{het}}$ and $\mathbf{A}_{\text{hom}}$) using the node features and extended node features respectively:

$$\mathbf{Q}_H = norm(\mathbf{V}_H w_{QH}),\ \mathbf{K}_H = norm(\mathbf{V}_H w_{KH}),\ \mathbf{A}_{\text{het}} = \mathbf{Q}_H\mathbf{K}_H^T/\sqrt{D'} \tag{2}$$

$$\mathbf{Q} = norm(\mathbf{V}w_Q),\ \mathbf{K} = norm(\mathbf{V}w_K),\ \mathbf{A}_{\text{hom}} = \mathbf{Q}\mathbf{K}^T/\sqrt{D} \tag{3}$$

where $w_{QH}, w_{KH} \in \mathbb{R}^{D'\times\frac{D'}{2}}, w_Q, w_K \in \mathbb{R}^{D\times\frac{D}{2}}$ are the learnable parameters and $D' = D + D_H$. The $norm()$ is a normalization operator.

**Attention Matrices Masking.** Subsequently, the heterogeneous components are extracted from the $\mathbf{A}_{\text{het}}$ and flattened into a base vector. The same process is applied to the homogeneous components of $\mathbf{A}_{\text{hom}}$.

$$B_{\text{het}} = \text{Flatten}(\mathbf{A}_{\text{het}}),\ for\ \mathbf{A}_{\text{het}_{ij}}\ where\ Type(v_i) \neq Type(v_j), \tag{4}$$

$$B_{\text{hom}} = \text{Flatten}(\mathbf{A}_{\text{hom}}),\ for\ \mathbf{A}_{\text{hom}_{ij}}\ where\ Type(v_i) = Type(v_j), \tag{5}$$

where $B_{het} \in \mathbb{R}^{P\times 1}$ and $B_{hom} \in \mathbb{R}^{(N^2-P)\times 1}$ are the base vectors and $P$ is the number of elements in the heterogeneous part. The base vector is passed through a linear layer to derive a threshold $\lambda_{het}$, which is used to compute the masked hetero-attention matrix $\mathbf{A}_{\text{het}}^{\text{masked}}$. The same procedure is applied to mask the homo-attention matrix. The detailed process is as follows:

$$\lambda_{het} = B_{\text{het}} w_{B1},\ \lambda_{hom} = B_{\text{hom}} w_{B2}, \tag{6}$$

$$\mathbf{A}_{\text{het}}^{\text{masked}} = \mathbf{A}_{\text{het}} \odot \text{Sign}(\mathbf{A}_{\text{het}} - \lambda_{het}),\ \mathbf{A}_{\text{hom}}^{\text{masked}} = \mathbf{A}_{\text{hom}} \odot \text{Sign}(\mathbf{A}_{\text{hom}} - \lambda_{hom}), \tag{7}$$

where $w_{B1} \in \mathbb{R}^{P\times 1}$, $w_{B2} \in \mathbb{R}^{(N^2-P)\times 1}$are the weight parameters and $\odot$ denote the Hadamard product. The $Sign()$ function returns 1 for positive inputs and 0 otherwise. Then, we combine $\mathbf{A}_{\text{het}}^{\text{masked}}$ and $\mathbf{A}_{\text{hom}}^{\text{masked}}$ to get the final heterogeneous masked attention matrix $\mathbf{A}^{\text{HM}}$:

$$A_{ij}^{\text{HM}} = \begin{cases} A_{\text{het}_{ij}}^{\text{masked}},\ Type(V_i) \neq Type(V_j) \\ A_{\text{hom}_{ij}}^{\text{masked}},\ Type(V_i) = Type(V_j) \end{cases} \tag{8}$$

### 2.3 Attention-Guided Path Convolution

Brain regions communicate through both direct and indirect pathways. A path $P$, can be seen as a sequence of $n$ contiguous edges connecting a starting node $v_i$ to a terminal node $v_j$. A 2-hop path $i-k-j$ is denoted as $p_{ikj}^2 = \{e_{ik}, e_{kj}\}$. If $i = k = j$, $p_{ikj}^0$ denotes a 0-hop path which is a self-loop. If $i = k \neq j$ or $i \neq k = j$, $p_{ikj}^1$ denotes 1-hop path which is a direct connection. For simplicity, we limit the maximum number of hops to 2. Thus, a set of paths between node $v_i$ and $v_j$ is represented as $P_{ij} = \{p_{ikj} | \forall k \in N\} = \{e_{ik}, e_{kj} | \forall k \in N\}$.

We utilized the same path convolution as [10], which can be represented as:

$$E'_{ij} = P_{ikj} w_k = (E_{ik} + E_{kj})\, w_k, \tag{9}$$

where $w_k$ is the weight parameter and $E'_{ij}$ is the learned edge feature. Additionally, we use masked attention $\mathbf{A}^{\text{HM}}$ to guide the convolution kernel to learn accurate features, which is shown in the green part of Fig.2 and is denoted as:

$$E'_{ij} = \left(A_{ik}^{HM} E_{ik} + A_{kj}^{HM} E_{kj}\right) w_k, \tag{10}$$

where $w_k$ is the weight parameter and $E'_{ij}$ is the learned edge feature. To mitigate the over-smoothing issue in graph-based methods, we adopt an SE block after path convolution. The final formulation can be represented as:

$$\mathbf{E}^{new} = SE(AGPConv(\mathbf{E}, \mathbf{A}_{ij}^{\text{HM}})), \tag{11}$$

where $E^{new}$ is the output edge feature, and $AGPConv()$ is the attention-guided path convolution we just introduced.

**Table 1.** Disease diagnosis results (mean±std) of comparison models and our method on ABIDE dataset.(**bold**:best; underline:runner-up)

| **Model** | $ACC\,(\%)\uparrow$ | $F1\,(\%)\uparrow$ | $SEN\,(\%)\uparrow$ | $SPE\,(\%)\uparrow$ |
|---|---|---|---|---|
| MLP | $71.73 \pm 3.78$ | $71.39 \pm 3.90$ | $67.55 \pm 8.81$ | $74.49 \pm 7.38$ |
| CNN | $67.93 \pm 2.67$ | $67.57 \pm 2.98$ | $64.89 \pm 13.13$ | $70.40 \pm 7.88$ |
| GCN | $67.24 \pm 2.07$ | $66.94 \pm 1.96$ | $61.05 \pm 4.91$ | $72.79 \pm 5.10$ |
| GBT | $72.77 \pm 3.42$ | $72.43 \pm 3.75$ | $70.36 \pm 8.34$ | $74.20 \pm 8.18$ |
| ALTER | $73.34 \pm 3.21$ | $72.89 \pm 3.52$ | $71.40 \pm 11.18$ | $74.19 \pm 10.03$ |
| BrainNetCNN | $71.96 \pm 3.65$ | $71.80 \pm 3.64$ | $68.56 \pm 8.12$ | $74.55 \pm 7.47$ |
| BC-GCN | $74.16 \pm 3.08$ | $73.95 \pm 2.96$ | $72.73 \pm 8.55$ | $74.90 \pm 9.30$ |
| PH-BTN | $\underline{74.62 \pm 3.81}$ | $\underline{74.58 \pm 3.81}$ | $\underline{73.83 \pm 6.01}$ | $\underline{75.22 \pm 6.51}$ |
| HM-AGPC (ours) | $\mathbf{76.46 \pm 3.16}$ | $\mathbf{76.38 \pm 3.34}$ | $\mathbf{76.15 \pm 8.39}$ | $\mathbf{76.54 \pm 5.17}$ |

### 2.4 Readout and Prediction

We employ the same Edge Pooling (EP) and Node Pooling (NP) techniques as [9]. The EP layers aggregate edge features into node embeddings, while the NP layers transform node embeddings into a final graph representation $G$. Graph representation $G$ is then passed through a multi-layer perceptron (MLP) to generate the final prediction $y$. Specifically, during inference, the final prediction for each subject is obtained by averaging $y$ across all its slices.

## 3 Experiment

**Data Processing and Experiment Setup.** We validate our HM-AGPC on the open-source dataset Autism Brain Imaging Data Exchange (ABIDE) [19].We follow the approach in [20], selecting a subset of 871 higher-quality samples (403ASD and 468HC) for the disease diagnosis task. We adopt AAL atlas ($N = 116$ ROIs) [21] and the brain functional partition (frontal, parietal, temporal, occipital, insular, cerebellum) as prior knowledge, where the number of categories $T = 6$. For each scan, we adopt a sliding window of 30 points and a stride of 20 points, cutting it into multiple slices to calculate pairwise Pearson's correlation of ROIs. We evaluate HM-AGPC and other related models (classic deep learning models - MLP, CNN and GCN; transformer-based graph models - GBT [17] and ALTER [16]; edge-based graph models - BrainnetCNN [9], BC-GCN [10] and PH-BTN [13]) through 10-fold cross-validation. We utilize ADAM optimizer with the learning rate of 0.0001, batch size of 64, and train models for 10 times with 60 epochs. Cross Entropy loss is used for the disease diagnosis task. We initialize weights with a standard normal distribution and train models on a GeForce GTX3090 with 24GB. For evaluation metrics, we choose classification accuracy (ACC), weighted F1-score (F1), sensitivity (SEN) and specificity (SPE). For all baselines, we strictly follow their experimental setup, fine-tuning hyperparameters on our data and reaching the best performance to ensure fairness. Our code is now open-sourced at: https://github.com/SCUT-Xinlab/HM-AGPC.

**Table 2.** Ablation study results (mean±std) of variants and our method on ABIDE dataset. Hetero-Prior Ablation refers to heterogeneous prior ablation; Left-right means dividing brain regions into left and right partitions; RandomX means randomly dividing brain regions into X categories. (**bold**:best)

| | **Model** | $ACC\,(\%)\uparrow$ | $F1\,(\%)\uparrow$ | $SEN\,(\%)\uparrow$ | $SPE\,(\%)\uparrow$ |
|---|---|---|---|---|---|
| Module Ablation | w/o Mask | $74.62 \pm 1.93$ | $74.07 \pm 2.58$ | $72.66 \pm 8.35$ | $75.34 \pm 7.86$ |
| | w/o heterogeneity | $75.20 \pm 2.47$ | $74.82 \pm 3.02$ | $72.58 \pm 6.85$ | $77.22 \pm 5.51$ |
| Hetero-Prior Ablation | left-right | $74.04 \pm 3.16$ | $73.81 \pm 3.17$ | $71.95 \pm 9.61$ | $75.80 \pm 10.17$ |
| | random4 | $74.26 \pm 4.21$ | $74.08 \pm 4.34$ | $73.59 \pm 5.40$ | $75.54 \pm 6.98$ |
| | random6 | $73.70 \pm 3.30$ | $73.51 \pm 3.46$ | $70.74 \pm 6.67$ | $75.82 \pm 9.26$ |
| | random8 | $74.50 \pm 3.35$ | $74.28 \pm 4.07$ | $66.65 \pm 7.64$ | $\mathbf{81.11 \pm 5.67}$ |
| Ours | HM-AGPC | $\mathbf{76.46 \pm 3.16}$ | $\mathbf{76.38 \pm 3.34}$ | $\mathbf{76.15 \pm 8.39}$ | $76.54 \pm 5.17$ |

**Performance Analysis.** As shown in Table 1, we demonstrate the results of the proposed HM-AGPC and related alternative methods on the ABIDE dataset. We can see that the proposed HM-AGPC outperformed alternative models on the disease diagnosis task with 76.46% ACC, 76.38% F1, 76.15% SEN and 76.54% SPE. Edge-based methods (BC-GCN, PH-BTN and proposed HM-AGPC) surpass node-based approaches (GBT and ALTER), indicating that the functional brain network is more suitable for modeling using edge features. Furthermore, the superior performance of HM-AGPC over BC-GCN and especially PH-BTN (a heterogeneous path-based algorithm) demonstrates that the heterogeneous masking mechanism provides reasonable positive guidance for path convolution, optimizing the convolution kernel without directly introducing additional parameters, thereby yielding more accurate diagnostic results.

**Ablation Study.** We conduct two ablation experiments: module ablation and heterogeneous prior ablation. In module ablation, we test two variants: w/o heterogeneity (pure homogeneous mask) and w/o Mask (pure attention guidance). As demonstrated in Table 2, both variants show performance declines. Without the precise characterization of heterogeneity for brain connectivity, the homogeneous mask fails to capture diverse brain region relationships, leading to a decline in results. In the absence of mask, the attention matrix loses its ability to filter out redundant interactions, thereby reducing its positive guiding effect on the convolution kernel. These results indicate that a proper masking strategy can effectively model the brain network while highlighting important connections, which is crucial for the attention-guided mechanism.

In the heterogeneous prior ablation study, we test four models with different priors: left-right, random4, random6, and random8. The left-right model divides the brain regions into two major categories (left and right), while the randomX models randomly partition the brain regions into X categories. Results in Table 2 show performance declines across all models, indicating that a well-designed brain prior partition is vital for the heterogeneous attention mask. It is also notable that although random8 outperformed random6 and random4 in accuracy,

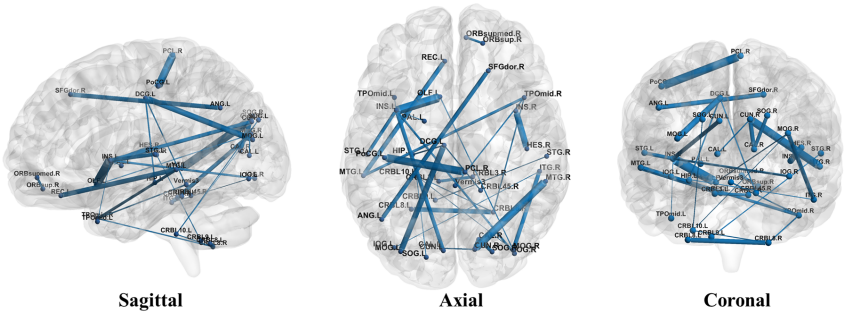


**Fig. 3.** The visualization results by the back-tracking method. We visualize the top-30 important functional brain connections learned by our model for the disease diagnosis task in ABIDE dataset.

the results for sensitivity and specificity were highly imbalanced. This suggests that more categories do not necessarily lead to better performance.

**Visualization.** We operate the gradient back-tracking method [10] to visualize the top-30 important brain functional connections learned by our model for the disease diagnosis task in ABIDE dataset. From Fig.3, it can be observed that the primary connections are concentrated among temporal lobe, occipital pole, insula and cerebellum. In particular, the insula demonstrates strong importance, which aligns with the findings in [22,23], suggesting the dysfunction of emotional and social processing for ASD patients. There are also many crucial connections in the occipital and temporal lobes, which may reflect limitations in visual and auditory information integration functions in children with ASD [24–26]. Additionally, connections in the cerebellar regions show strong significance, which may be related to motor coordination and cognitive function regulation. Studies in [27] and [28] have pointed out that abnormal connectivity patterns in the cerebellum may affect motor control and executive functions in ASD patients. In summary, these key connectivity patterns not only reveal potential neural mechanisms of ASD but also demonstrate the advantages of our approach in terms of interpretability, providing new insights for the search of biomarkers.

## 4 Conclusion

In this paper, we propose a novel Heterogeneous Masked Attention-Guided Path Convolution (HM-AGPC) for functional brain network analysis. HM-AGPC introduces a heterogeneous masked attention mechanism to emphasize valuable heterogeneous relationships while minimizing redundant interactions, enhancing robustness and interpretability. The framework also employs an attention-guided path convolution strategy, which utilizes attention weights to guide convolution kernels toward the most salient features and pathways. With the support of these

two key mechanisms, HM-AGPC demonstrates exceptional performance in the disease diagnosis task. Moreover, HM-AGPC provides a novel approach for exploring heterogeneous brain connectivity relationships, offering new insights for brain disorder analysis and the discovery of related biomarkers.

**Acknowledgments.** This study was funded by General Program of the National Natural Science Foundation of China (NSFC-62471185), Guangdong Basic and Applied Basic Research Foundation (2024A1515010180), Key-Area Research and Development Program of Guangdong Province (2023B0303040001) and the Guangdong Provincial Key Laboratory of Human Digital Twin (2022B1212010004).

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

## References

1. Dosenbach, N.U.F., et al.: Prediction of individual brain maturity using fMRI. Science **329**(5997), 1358–1361 (2010).
2. Nakai, T. and Nishimoto, S.: Quantitative models reveal the organization of diverse cognitive functions in the brain. Nat. Commun. **11**(1), 1142 (2020).
3. Santana, C. P., et al.: Rs-fMRI and machine learning for ASD diagnosis: a systematic review and meta-analysis. Sci. Rep. **12**(1), 6030 (2022)
4. Liu, M., Li, B., and Hu., D.: Autism spectrum disorder studies using fMRI data and machine learning: a review. Front. Neurosci. **15** (2021)
5. Lee, B., Kang, U., Chang, H., Cho, K.H.: The hidden community architecture of human brain networks. Sci. Rep. **12**(1), 3540 (2022)
6. Wein, S., et al.: A graph neural network framework for causal inference in brain networks. Sci. Rep. **11**(1), 8061 (2021)
7. Li, X., et al.: BrainGNN: interpretable brain graph neural network for fMRI analysis. Med. Image Anal. **74**, 102233 (2021)
8. Dong, Z., et al.:: Beyond the snapshot: brain tokenized graph transformer for longitudinal brain functional connectome embedding. In: Greenspan, H., et al. MICCAI 2023, LNCS, vol 14224, pp. 328–337. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-43904-9_34
9. Kawahara, J., et al.: BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage **146**, 1038–1049 (2017)
10. Li, Y., et al.: Brain Connectivity based graph convolutional networks and its application to infant age prediction. IEEE Trans. Med. Imaging **41**(10), 2764–2776 (2022)
11. Grachev, I.D., Apkarian, A.V.: Chemical Heterogeneity of the Living Human Brain: A Proton MR Spectroscopy Study on the Effects of Sex, Age, and Brain Region. NeuroImage **11**(5), 554–563 (2000)
12. Haak, K.V., Beckmann, C.F.: Understanding brain organisation in the face of functional heterogeneity and functional multiplicity. NeuroImage **220**, 117061 (2020)
13. Fang, R., et al.: Path-based heterogeneous brain transformer network for resting-state functional connectivity analysis. In: Greenspan, H., et al. MICCAI 2023, LNCS, vol 14227, pp. 328–337. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-43993-3_32

14. Xia, J., Chan, Y. H., Girish, D., Rajapakse, J. C.: IMG-GCN: Interpretable modularity-guided structure-function interactions learning for brain cognition and disorder analysis. In: Linguraru, M.G., et al. MICCAI 2024, LNCS, vol 15010, pp. 470–480. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72117-5_44
15. Kan, X., Dai, W., Cui, H., Zhang, Z., Guo, Y., Yang, C.: Brain network transformer. In: Advances in Neural Information Processing Systems, vol. 35, pp. 25586–25599 (2022)
16. Yu, S., Jin, S., Li, M., Sarwar, T.and Xia, F.: Long-range brain graph transformer. In: Advances in Neural Information Processing Systems, vol. 37, pp. 24472–24495 (2025)
17. Peng, Z., He, Z., Jiang, Y., Wang, P., Yuan, Y.: GBT: Geometric-oriented brain transformer for autism diagnosis. In: Linguraru, M.G., et al. MICCAI 2024, LNCS, vol. 15012, pp. 142–152. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72390-2_14
18. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
19. Cameron, C., et al.: The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives. Front. Neuroinform. **7** (2013)
20. Abraham, A., et al.: Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. NeuroImage **147**, 736–745 (2017)
21. Tzourio-Mazoyer, N., et al.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject Brain. NeuroImage **15**(1), pp. 273–289 (2002)
22. Nomi, J. S., Molnar-Szakacs, I., Uddin, L. Q.: Insular function in autism: update and future directions in neuroimaging and interventions. Prog. Neuropsychopharmacol. Biol. Psychiatry. **89**, 412–426 (2019)
23. Caria, A., de Falco, S.: Anterior insular cortex regulation in autism spectrum disorders. Front. Behav. Neurosci. **9** (2015)
24. Uddin, L. Q., et al.: Salience network–based classification and prediction of symptom severity in children with autism. JAMA Psychiat. **70**(8), 869–879 (2013)
25. Zilbovicius, M., et al.: Temporal lobe dysfunction in childhood autism: a PET study. Am. J. Psychiat. **157**(12), 1988–1993 (2000)
26. Jung, M., et al.: Decreased structural connectivity and resting-state brain activity in the lateral occipital cortex is associated with social communication deficits in boys with autism spectrum disorder. NeuroImage **190**, 205–212 (2019)
27. Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., Cauraugh, J. H.: Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. J. Autism. Dev. Disord. **40**(10), 1227–1240 (2010)
28. Minshew, N. J., Sung, K., Jones, B. L., Furman, J. M.: Underdevelopment of the postural control system in autism. Neurology **63**(11), 2056–2061 (2004)