Towards Graph Neural Networks with Domain-Generalizable Explainability for fMRI-Based Brain Disorder Diagnosis

Xinmei Qiu[1], Fan Wang[2,4,✉], Yongheng Sun[1], Chunfeng Lian[1,3,✉], and Jianhua Ma[2,3,✉]

[1] School of Mathematics and Statistics, Xian Jiaotong University, Xi'an, China
chunfeng.lian@xjtu.edu.cn
[2] The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China
{fan.wang,jhma}@xjtu.edu.cn
[3] Pazhou Lab (Huangpu), Guangzhou, China
[4] The First Affliated Hospital of Xi'an Jiao Tong University, Xi'an, China

Abstract. Graph neural networks (GNNs) represent a cutting-edge methodology in diagnosing brain disorders via fMRI data. Explainability and generalizability are two critical issues of GNNs for fMRI-based diagnoses, considering the high complexity of functional brain networks and the strong variations in fMRI data across different clinical centers. Although there have been many studies on GNNs' explainability and generalizability, yet few have addressed both aspects simultaneously. In this paper, we unify these two issues and revisit the domain generalization (DG) of fMRI-based diagnoses from the view of explainability. That is, we aim to learn domain-generalizable explanation factors to enhance center-agnostic graph representation learning and therefore brain disorder diagnoses. To this end, a specialized meta-learning framework coupled with explainability-generalizable (XG) regularizations is designed to learn diagnostic GNN models (termed XG-GNN) from fMRI BOLD signals. Our XG-GNN features the ability to build nonlinear functional networks in a task-oriented fashion. More importantly, the group-wise differences of such learned individual networks can be stably captured and maintained to unseen fMRI centers to jointly boost the DG of diagnostic explainability and accuracy. Experimental results on the ABIDE dataset demonstrate the effectiveness of our XG-GNN. The source code will be released on https://github.com/ladderlab-xjtu/XG-GNN.

Keywords: Domain generalization · Explainability · Meta-learning · Graph neural networks · fMRI

1 Introduction

Brain network analysis, especially through the use of functional magnetic resonance imaging (fMRI) techniques, plays a crucial role in understanding neurological developments and degenerations, along with their associated disorders [16].

For instance, translational psychiatry research has shown that the functional changes and remodeling associated with Autism spectrum disorders (ASD) may appear before the behavioral symptoms [5], highlighting the clinical importance of fMRI for early diagnosis and intervention. Numerous studies in the literature have aimed to develop classification models using fMRI data to identify biomarkers and diagnose neuropsychiatric disorders [9,12]. Among these, graph neural networks (GNNs) have garnered significant attention, which stems from their specialized learning mechanisms well-suited for the processing of data structured as graphs [11,15,7]. Overall, the practical usage of these GNN-based diagnostic methods heavily relies on their *explainability* and *generalization*. This is particularly challenging given the high complexity of functional brain networks and the strong variations in fMRI data across different clinical sites.

In recent years, the push to enhance the explanation and generalization capabilities of GNNs has gained momentum, yet few efforts have addressed both aspects simultaneously. For instance, to make GNN-based diagnostics more transparent, [14] and [2] design learnable graph-pooling and graph-masking operations to identify subject-level and group-level explanations from the functional connectivity (FC) network defined by linear correlations, respectively. In [20], the authors integrated modularity prior of the human brain into DNNs to enhance the explainability of diagnoses through dynamic linear FC. It is important to note that linear FCs, such as those based on Pearson correlations, overlook the temporal order and struggle to fully capture the complexities of brain networks. In response, studies like [8] and [23] have developed GNNs that learn nonlinear FCs from fMRI BOLD signals in a task-oriented manner. These fully learnable methods excel in identifying abnormal brain connections tied to particular disorders, providing more explicit explanations for biomarker discovery and diagnosis. However, much existing research on explainability fails to consider the significant domain shifts in fMRI data, which severely limits the ability of diagnostic models and identified biomarkers to be generalized across various clinical centers. To improve generalization in diagnosing brain disorders, researchers have explored domain adaptation (DA) [13] and domain generalization (DG) [12]. For example, [12] introduced a meta-learning framework to develop site-invariant (i.e., DG) models for ASD diagnosis using linear FCs. It is worth mentioning that DG does not require fine-tuning on the unknown target set, making it more applicable than DA in the context of medical imaging. However, these studies primarily concentrate on improving the generalization of classification outcomes rather than the explanatory power. It limits the models' ability to accurately classify because identifying generalizable biomarkers is a fundamental step towards achieving reliable diagnoses of complex brain disorders.

Consistent with the perspectives shared in our previous work [22], we propose that the explicit learning of explanatory factors to enhance discriminative representation learning represents a logical strategy for concurrently ensuring reliable diagnostic outcomes and detailed explainability. This is particularly relevant given the inherent complexity, noise, and redundancy typical of functional brain networks. *Diverging from this, our study here goes a step*

further by unifying explainability with generalizability. That is, we aim to learn domain-generalizable explanation factors to enhance domain-agnostic extraction of discriminative representations. This, in turn, facilitates accurate and explainable diagnoses across various clinical centers. To achieve this, an explainability-generalizable GNN (termed XG-GNN) is developed to learn task-oriented brain networks and associated diagnoses from the BOLD signals of pre-parcellated brain regions. We employ a specialized meta-learning framework to train such an end-to-end network, focusing on the dual goals of explanability and diagnostic accuracy. Specifically, in the outer-loop episode of the meta-learning procedure, we define explainability-generalizable (XG) regularizations to maintain center-agnostic group-level differences of learned brain networks, based on which the inner-loop episode naturally simulates scenarios where these consistently identified, detailed explanatory factors enhance the learning of fine-grained discriminative representations, leading to precise diagnoses across unseen domains. Experimental evaluations on the public ABIDE benchmark indicate that our XG-GNN leads to state-of-the-art DG performance in fMRI-based ASD diagnosis with verifiable explanations.

Overall, the main contributions of the paper are threefold:

1. To the best of our knowledge, this is the first attempt to unify explainability and generalizability in the task of brain disorder diagnosis across multi-center fMRI data. Learning domain-agnostic explanatory factors to enhance discriminative feature representation is an intuitive strategy to achieve reliable DG of explainable diagnostic models in such a challenging task.
2. We introduce a specialized meta-learning framework, augmented with explicit XG regularizations, to enable the simultaneous DG of both diagnostic explanations and outcomes.
3. Our XG-GNN learns individualized brain networks while also identifying their class/group-wise differences across different centers, featuring its capability to concurrently recognize subject-specific and group-consistent abnormalities related to particular disorders.

2 XG-GNN

2.1 Architecture

Our XG-GNN performs explainable diagnosis by using as input the BOLD signals of the brain regions of interests (ROIs), e.g., pre-parcellated according to the CC200 atlas [1]. As shown in Fig. 1, the model consists of two main components, i.e., a brain-graph learner based on multi-head self-attention mechanisms (MHSA) and a diagnoser based on graph convolutional network (GCN) [10] blocks. In an end-to-end learnable fashion, the MHSA-based graph learner builds nonlinear functional networks, based on which the GCN-based diagnoser outputs the respective disease status, e.g., ASD or typical development (TD).

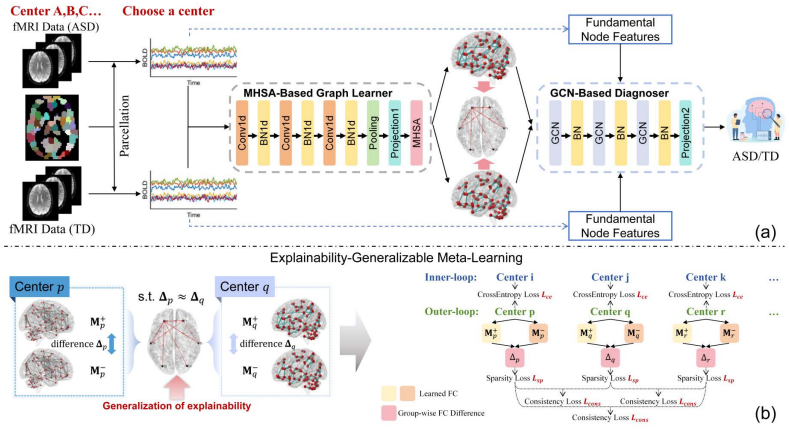

Fig. 1: Illustration of our XG-GNN for multi-site brain disorder diagnosis.

1) MHSA-Based Graph Learner: Let $\mathbf{X} \in \mathbb{R}^{n \times T}$ be the raw fMRI BOLD signals of a subject, where n is the number of ROIs and T represents the length of each ROI's time series. Then, by regarding the ROIs as the vertices, say $V = \{v_1, \ldots, v_n\}$, the goal is to learn an undirected weighted graph $G = (V, \mathcal{E}, \mathbf{M})$ in terms of $\mathbf{X}$, where $\mathcal{E}$ represents the set of edges, namely the collection of connected vertices (v_i, v_j) from v_i to v_j, and $\mathbf{M} \in \mathbb{R}^{n \times n}$ is the learned nonlinear connectivity matrix among the ROIs. Such a graph G is built in a task-oriented fashion to capture subject-specific disorder patterns.

To this end, we design a simple but effective sub-network, i.e., MHSA-based graph learner, with fundamental blocks. As shown in Fig. 1 (a), it first applies a set of three consecutive one-dimensional convolutional (Conv1d) layers followed by batch normalization (BN1d) and rectified linear unit (ReLU) operations to map each ROI's BOLD into a nonlinear feature space, such as

$$\mathbf{F}_i = \text{BN1d}_i(\text{Conv1d}_i(\mathbf{X})) = \text{BN1d}_i(\mathbf{X}\mathbf{W}_i + \mathbf{b}_i), \tag{1}$$

where $\mathbf{F}_i \in \mathbb{R}^{n \times h_i}$ is the output BOLD embedding from the ith layer (parameterized by $\mathbf{W}_i$ and $\mathbf{b}_i$), and h_i is the respective number of channels. Then, after squeezing the BOLD embedding from the last Conv1d layer with a channel-wise max pooling operation, the projection module applies two linear transformations followed by a softmax activation to map the features to the desired output size, the resulting tensor $\mathbf{f} \in \mathbb{R}^{n \times h}$ is further processed by a MHSA block to capture the complex nonlinear associations between different brain regions. In MHSA, it first performs multi-head linear transformations to produce the query ($\mathbf{Q}$), key ($\mathbf{K}$), and value ($\mathbf{V}$) from $\mathbf{f}$ for each head, such as $\mathbf{Q} = \mathbf{W}_q\mathbf{f} + \mathbf{b}_q, \mathbf{K} = \mathbf{W}_k\mathbf{f} + \mathbf{b}_k, \mathbf{V} = \mathbf{W}_v\mathbf{f} + \mathbf{b}_v$, respectively. After that, each ROI's embedding is updated by aggregating the cross-ROI associations:

$$\text{MHSA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{2}$$

where d_k is the channel dimension of $\mathbf{K}$. Finally, the outputs of all heads are merged to produce the final BOLD embedding $\mathbf{A} \in \mathbb{R}^{n \times m}$, based on which the brain connectivity is defined as $\mathbf{M} = \mathbf{A}\mathbf{A}^T$.

2) GCN-Based Diagnoser: Given the brain graph $G = (V, \mathcal{E}, \mathbf{M})$ with $\mathbf{M}$ learned by the MHSA-based graph learner, we further develop a fundamental GCN-based classifier to conduct diagnosis. There are three GCN blocks in such a diagnoser, with each containing graph convolution followed by BN and LeakyReLU activation. Specifically, given input node features $\mathbf{g}_{i-1} \in \mathbb{R}^{n \times r_{i-1}}$, each block outputs the updated features $\mathbf{g}_i \in \mathbb{R}^{n \times r_i}$ (parameterized by $\mathbf{U}_i$) as

$$\mathbf{g}_i = \mathrm{BN}(\mathrm{GCN}(\mathbf{M}, \mathbf{g}_{i-1})) = \mathrm{BN}(\mathrm{RELU}(\mathbf{M}\mathbf{g}_{i-1}\mathbf{U}_i)). \tag{3}$$

The node features input into the first block are simply initialized by cross-ROI Pearson correlation coefficients. Finally, the node-level feature representations are flattened to produce graph-level diagnostic outcomes with a projection consisting of multiple linear layers and activation functions.

2.2 Explainability-Generalizable Meta-Learning

1) Explainability-Generalizable (XG) Regularization: Our model learns nonlinear FC matrices and associated diagnoses in a task-oriented fashion, with the FC naturally capturing subject-specific patterns tied to the diagnostic outcome. To further enhance the group-wise explanations as well as their generalization across heterogeneous sites, we design complementary XG regularizations by leveraging fundamental prior regarding the early status of neuropsychiatric disorders (e.g., ASD), independent of fMRI centers.

Sparsity of Inter-Group FC Differences. Considering that the functional abnormalities related to the early disorder stage are typically regionalized or not widely diffused over the whole brain, it is straightforward to assume that the group-wise differences between early disorder and health in terms of FCs are relatively sparse [21]. To enhance such group-wise explainability of learned FCs, we design a targeted entropy-based regularization L_{sp}. Specifically, let $D_1, ..., D_K$ be K different clinical sites. For each site D_k, there are two different subject groups, i.e., patients and healthy controls. We average the learned FCs in each subject group of D_k, denoted as $\overline{\mathbf{M}^+_{D_k}}$ and $\overline{\mathbf{M}^-_{D_k}}$, respectively. Then, we quantify the group-wise FC differences as $\mathbf{\Delta}_{D_k} = \left|\overline{\mathbf{M}^+_{D_k}} - \overline{\mathbf{M}^-_{D_k}}\right| \in \mathbb{R}^{n \times n}$. To enhance the sparsity of such group-wise FC differences, we define the penalty as

$$L_{\mathrm{sp}} = -\| \sum_k \mathbf{\Delta}_{D_k} log \mathbf{\Delta}_{D_k} \|_2. \tag{4}$$

Cross-Site Consistency of Inter-Group FC Differences. Considering that, given a particular disorder, the inter-group differences of FCs are typically stable and independent of fMRI data sites [19], we design a cross-site consistency regularization to enhance such center-agnostic group-wise explanations, such as

$$L_{\mathrm{cons}} = -\sum_{i,j} \| \frac{\left|\mathbf{\Delta}_{D_i} \cdot \mathbf{\Delta}_{D_j}\right|}{\left|\mathbf{\Delta}_{D_i}\right| \left|\mathbf{\Delta}_{D_j}\right|} \|_2, \tag{5}$$

Algorithm 1 Explainability-Generalizable Meta-Learning

Require: Source domain dataset S;
Meta-parameters: step size γ, number of meta-iterations K
Ensure: Updated XG-GNN model parameters θ
1: Randomly initialize model parameters θ
2: **for** $k = 1$ **to** K **do**
3: Randomly partition S into inner-loop set S_{inner} and outer-loop set S_{outer};
4: Inner-loop: Update θ'_k on S_{inner} using cross-entropy loss $\theta'_k = \theta_{k-1} - \gamma \nabla_{\theta_{k-1}} L_{\text{ce}}(S_{\text{inner}}; \theta_{k-1})$;
5: Outer-loop: Update XG-GNN model parameters θ on S_{outer} using meta loss $\theta_k = \theta_{k-1} - \gamma \nabla_{\theta'_k} L_{\text{meta}}(S_{\text{outer}}; \theta'_k)$.
6: **end for**

where i and j represent any two different centers. Overall, the combination of Eqs. (4) and (5) jointly form the XG regularization on learning domain-generalizable, group-wise explanations to boost diagnoses.

2) Bi-Level Meta-Learning Algorithm: To achieve joint DG of explanations and diagnoses, we design a bi-level meta-learning algorithm to train our GNN model under the XG regularization, such as shown in Algorithm 1. Given the fMRI data from multiple source-domain centers, i.e., $S = \{D_1, ..., D_K\}$, the algorithm simulates the DG scenario to enhance the stability of learned explanation factors and associated diagnostic outcomes when applied to potentially unseen (i.e., target-domain) centers. As shown in Fig. 1 (b), in the inner-loop of the bi-level optimization, we randomly sample a few subsets of source-domain centers (say S_{inner}), on which the GNN model is trained by simply L_{ce} for classification. Specifically in the outer-loop of the bi-level optimization, we further sample the remaining source-domain centers (say S_{outer}), on which the GNN model is trained by minimizing a meta-learning loss:

$$L_{\text{meta}} = L_{\text{ce}} + \alpha L_{\text{sp}} + \beta L_{\text{cons}}, \tag{6}$$

where L_{sp} and L_{cons} are from Eqs. (4) and (5), respectively, α and β are tuning parameters, and L_{ce} is a general cross-entropy loss for classification. Due to the implicit function relationship established by bi-level optimization [4], such a meta-learning algorithm effectively drives the group-wise explanation factors stably captured in the outer-loop to be generalized to boost diagnoses in the inner-loop (i.e., simulated DG situation). For more details regarding the meta-learning procedure and Algorithm 1, please refer to the *Supplementary Material*.

3 Experiments

3.1 Experimental Setup

1) Dataset: We evaluated our XG-GNN on the publicly accessible ABIDE (Autism Brain Imaging Data Exchange) dataset [3]. Specifically, the resting-state fMRI data from 16 international imaging centers were used, with 416 ASD and 418 TD individuals, respectively. We used the CC200 atlas [1] to parcellate each brain and averaged the BLOD signals within each brain ROI.

Table 1: Diagnostic results (mean±std) for different competing methods on the three target domains from the ABIDE dataset.

Method	ACC(%)	AUC(%)	F1 Score(%)
BrainnetCNN [9]	65.35 ± 2.20	70.87 ± 1.62	62.04 ± 1.54
IBGNN [2]	65.26 ± 1.61	72.88 ± 2.41	62.01 ± 2.93
Coral[17]	65.46 ± 2.22	70.75 ± 1.99	64.97 ± 4.31
GenM [12]	62.98 ± 4.79	69.40 ± 6.51	57.22 ± 9.31
DuMeta[18]	68.78 ± 5.02	73.41 ± 2.78	69.44 ± 5.33
XG-GNN **(ours)**	$\mathbf{70.66 \pm 1.07}$	$\mathbf{76.32 \pm 0.55}$	$\mathbf{69.78 \pm 3.82}$

Table 2: Ablation studies regarding each key component of our XG-GNN.

Ablation experiment	Baseline	w/o XG	w/o L_{cons}	w/o L_{sp}	XG-GNN **(ours)**
ACC(%)	64.66	68.80	69.97	69.04	**70.66**
AUC(%)	70.27	75.57	**77.47**	75.94	76.32

2) Implementation Details: Our XG-GNN was implemented in PyTorch and was trained on a PC with one NVIDIA RTX 3060 GPU. The model was trained in the bi-level meta-learning framework by using the Adam optimizer for $5,000$ iterations. In both the inner-loop and outer-loops episodes, the learning rate was initally set as 0.001, which was decayed by half after every $1,000$ iterations. We randomly selected one center as the target domain and regarding the remaining 15 centers as the source domians, under the requirement that the number of subject in the target domain should be relatively large (e.g., > 95) for a reliable quantification of the domain generalization performance. The procedure was repeated 3 times, i.e., the training was executed 3 different times, and the outcomes on the 3 different target domains were averaged as the final results.

3) Competing Methods: Our XG-GNN was compared with two representative methods for fMRI-based classification, i.e., **BrainnetCNN** [9] and **IBGNN** [2]. It was also compared with three representative DG methods, including **GenM** [12], **Coral** [17], and **DuMeta** [18]. Notably, we used the original version of GenM, as it was previously developed for DG of fMRI-based diagnosis. Since Coral and DuMeta were not originally developed for fMRI, we adapted them to use the same GNN architecture as our method for fair comparisons.

4) Evaluation Metrics: To comprehensively evaluate the diagnostic performance, we used three complementary metrics, i.e., accuracy (**ACC**), balanced **F1-score**, and area under the receiver operating characteristic curve (**AUC**).

3.2 Results

1) Classification Results: The diagnostic results obtained by all competing methods on the three target domains are summarized in Table 1, from which we can draw at least two key observations. *First*, compared with BrainnetCNN and IBGNN, most DG methods (i.e., Coral, DuMeta, and our XG-GNN) led to better performance in terms of most metrics on the three unseen domains. It suggests that domain gaps do exist across different fMRI data centers, and DG plays a important role in enhancing the generalization of diagnostic models.

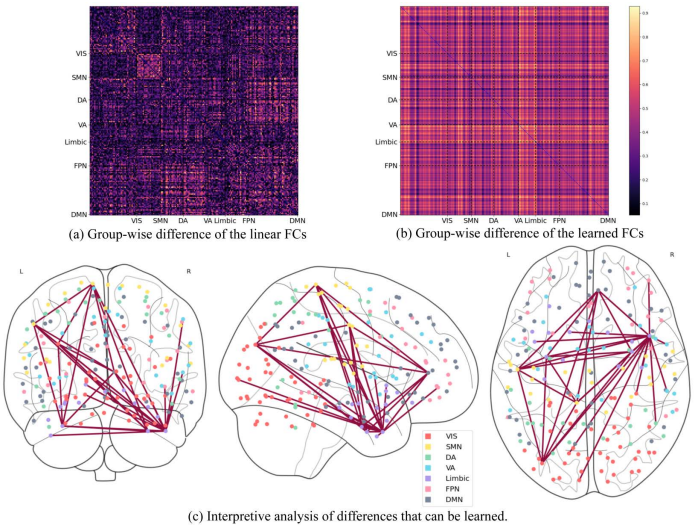

Fig. 2: Explanation results on ABIDE dataset.

Second, compared with all other methods, our XG-GNN consistently obtained significantly better results in terms of all metrics, implying the efficacy of our methodological designs in the task of fMRI-based brain disorder diagnosis. The variance of evaluation metrics obtained by our XG-GNN is much smaller compared to other methods, indicating its reproducibility and robustness.

2) Explanation Results: To check the efficacy of our XG-GNN in identifying domain-generalizable explanation factors, on the unseen target domains, we calculated the group-wise difference of the learned FCs and compared it with that quantified from linear FCs in terms of Pearson correlations. The corresponding visualization results are shown in Fig 2. From Fig. 2 (a), we can see that the group-wise difference in terms of the linear FCs is very uninformative (with almost all values be small). This suggests that the information captured within the brain network via Pearson correlation may exhibit minimal variance, thereby potentially limiting its utility in disorder diagnosis tasks. In contrast, as shown in Fig. 2 (b), the group-wise difference of the task-oriented FCs learned by our XG-GNN distinctly highlights the limbic modules as well as their connections with other network modules (highlighted in Fig. 2 (c)). This observation aligns consistently with the findings from previous neuroscience studies [6] that have reported significant differences in the limbic system between ASD and TD. It is worth mentioning that our model maintains this interpretability across different sites. These visualization results suggest that our method can stably capture group-wise connectivity abnormalities independent of fMRI centers, which is critical for the generalization of diagnostic explainablity and therefore accuracy.

3) Ablation Studies: We conducted ablation studies to verify the contributions of two key components of our method, i.e., the meta-learning strategy and the associated XG regularization. Let **Baseline** be our network trained without

meta-learning, and **w/o XG**, **w/o** $L_{\mathbf{cons}}$, and **w/o** $L_{\mathbf{sp}}$ denote our network trained in the meta-learning framework while without using the XG regularizations, L_{cons}, and L_{sp}, respectively. The comparisons between the original XG-GNN and these variants are summarized in Table 2. According to the results of **Baseline** and **w/o XG**, we can see both the meta-learning framework and the XG regularization (i.e., the combination of L_{cons} and L_{sp}) contributed to significant improvements of the diagnostic performance. On the other hand, according to the results of **w/o** $L_{\mathbf{cons}}$ and **w/o** $L_{\mathbf{sp}}$, we can see that the removing of any of these two terms resulted in the drop of performance, implying their complementarity in enhancing domain-generalizable explainability and diagnoses.

4 Conclusion

In this paper, we have proposed an explainability-generalizable GNN for the domain generalization of brain disorder diagnosis across multi-center fMRI data. To this end, a meta-learning framework integrating specialized regularizations have been developed to learn task-oriented brain networks that capture center-agnostic explanation factors to enhance discriminative graph representation learning and diagnostic outcomes. To the best of our knowledge, this is the first attempt to unify explainability and generalizability in the task of fMRI-based diagnosis. Experimental results on the ABIDE dataset have verified the efficacy of our method from both the aspects of diagnostic accuracy and explainability.

Acknowledgments. This work was supported in part by STI 2030-Major Projects (No. 2022ZD0209000), and NSFC Grants (Nos. 62101430, & 62101431).

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

References

1. Craddock, R.C., et al.: A whole brain fmri atlas generated via spatially constrained spectral clustering. Human Brain Mapping **33**(8), 1914–1928 (2012)
2. Cui, H., et al.: Interpretable graph neural networks for connectome-based brain disorder analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 375–385. Springer (2022)
3. Di Martino, A., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry **19**, 659–667 (2014)
4. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning. pp. 1126–1135. PMLR (2017)
5. Hashem, S., et al.: Genetics of structural and functional brain changes in autism spectrum disorder. Translational Psychiatry **10**(1), 229 (2020)
6. Haznedar, M.M., et al.: Limbic circuitry in patients with autism spectrum disorders studied with positron emission tomography and magnetic resonance imaging. American Journal of Psychiatry **157**(12), 1994–2001 (2000)
7. Kan, X., et al.: Brain network transformer. Advances in Neural Information Processing Systems **35**, 25586–25599 (2022)

8. Kan, X., et al.: Fbnetgen: Task-aware gnn-based fmri analysis via functional brain network generation. In: International Conference on Medical Imaging with Deep Learning. pp. 618–637. PMLR (2022)
9. Kawahara, J., et al.: Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage **146**, 1038–1049 (2017)
10. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (2017)
11. Ktena, S.I., et al.: Distance metric learning using graph convolutional networks: Application to functional brain networks. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20. pp. 469–477. Springer (2017)
12. Lee, J., et al.: Meta-modulation network for domain generalization in multi-site fmri classification. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24. pp. 500–509. Springer (2021)
13. Li, X., et al.: Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: Abide results. Medical Image Analysis **65**, 101765 (2020)
14. Li, X., et al.: Pooling regularized graph neural network for fmri biomarker analysis. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23. pp. 625–635. Springer (2020)
15. Li, X., et al.: Braingnn: Interpretable brain graph neural network for fmri analysis. Medical Image Analysis **74**, 102233 (2021)
16. Smith, S.M.: The future of fmri connectivity. Neuroimage **62**(2), 1257–1266 (2012)
17. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. In: Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14. pp. 443–450. Springer (2016)
18. Sun, Y., et al.: Dual meta-learning with longitudinally generalized regularization for one-shot brain tissue segmentation across the human lifespan. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 21061–21071. IEEE (2023)
19. Supekar, K., et al.: Robust, generalizable, and interpretable artificial intelligence–derived brain fingerprints of autism and social communication symptom severity. Biological Psychiatry **92**(8), 643–653 (2022)
20. Wang, Q., et al.: Modularity-constrained dynamic representation learning for interpretable brain disorder analysis with functional mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 46–56. Springer (2023)
21. Wang, Y., et al.: Social brain network of children with autism spectrum disorder: characterization of functional connectivity and potential association with stereotyped behavior. Brain Sciences **13**(2), 280 (2023)
22. Xue, C., et al.: Neuroexplainer: Fine-grained attention decoding to uncover cortical development patterns of preterm infants. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 202–211. Springer (2023)
23. Zhang, Y., Huang, H.: New graph-blind convolutional network for brain connectome data analysis. In: Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26. pp. 669–681. Springer (2019)