Graph Disentanglement Learning for fMRI Analysis: Decoupling Disease, Covariates, and Individual Variability

Shengjie Zhang[1,2,3], Zhuangzhuang Jiang[4], Xin Shen[5], Ziqi Yu[1,2,3], Xiang Chen[6], Xiao-Yong Zhang[1,2,3](✉), and Yuan Zhou[4](✉)

[1] Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, China.
[2] Faculty of Medical Imaging Technology, College of Health Science and Technology, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, China.
[3] Shanghai Key Laboratory of Child Brain and Development, Shanghai, 200127, China. zhangxiaoyong@sjtu.edu.cn
[4] School of Data Science, Fudan University, Shanghai, China. yuanzhou@fudan.edu.cn
[5] School of Mathematical Sciences, Beijing Normal University, Beijing, China.
[6] Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai, China.
The first two authors contribute equally to this work.

Abstract. Functional magnetic resonance imaging (fMRI) is a powerful tool for diagnosing neurological disorders. However, accurately distinguishing disease-related features from confounding covariates (e.g., age, gender, site) and individual variability remains a challenge. To tackle this problem, we propose a novel graph disentanglement learning (GDL) framework that decomposes the latent features from fMRI images into 3 components: disease-related features, covariate-related features, and individual variations. The covariate-related features are learned by aligning 2 subject similarity matrices between the features and the true covariates. The disease-related features are guided by a classification loss. We validate our method on 3 fMRI datasets: ADHD-200, schizophrenia (SCZ), and Presbycusis. The method outperforms existing approaches by an average of 0.5%, 1.7%, and 2.1% in accuracy on the 3 datasets respectively. Ablation studies confirm that our model is robust to hyperparameter selection. The disease-associated regions identified by our model align with established clinical findings. These results suggest that GDL is a promising tool for fMRI-based disease diagnosis and biomarker discovery. The code is publicly available at https://github.com/perpetualmachine/GDL_MICCAI.

Keywords: graph neural network · disentangled representation learning· disease diagnosis · fMRI

1 Introduction

Neurological disorders, such as attention deficit hyperactivity disorder (ADHD) and schizophrenia (SCZ), have gained significant attention due to their increasing contemporary prevalence [1]. Early diagnosis of neurological disorders can provide valuable information to subsequent treatments. One way of diagnosing neurological disorders is to use functional magnetic resonance imaging (fMRI) which measures blood oxygen level-dependent (BOLD) signals in the brain [2].

Existing methods for diagnosing neurological disorders using fMRI can be broadly categorized into traditional machine learning (ML) and deep learning (DL) approaches. Traditional ML methods typically involve a pipeline of feature extraction, e.g., amplitude of low frequency fluctuations (ALFF) [3] and regional homogeneity (ReHO) [4], followed by feature classification. These methods suffer from the limited features extracted from the traditional feature extraction techniques. In contrast, DL methods, particularly graph neural networks (GNNs) [5], automatically extract features from fMRI data, potentially offering better performance in disease diagnosis.

GNN-based fMRI analysis can be broadly categorized into graph-level classification and node-level classification [1]. Graph-level classification treats a graph as an instance and tries to classify the graphs [6]. Node-level classification combines all the graphs to a large population graph with the node being the individual graph and performs node classification [7]. In the learning process, various techniques, such as adversarial learning [8] and self-supervised learning (SSL) [9], have been adopted. Despite a plethora of these GNN-based fMRI methods, they face a critical challenge: they struggle to disentangle disease from covariates (e.g. age, gender, site, etc) and individual variability in feature extraction.

Recently, a few works have attempted to disentangle ages in a regression setting [10]. However, they usually only consider the age as the covariate and assume a linear mapping from features to ages. In contrast, real-world images are influenced by multiple covariates (e.g. age, gender, site), and enforcing a linear relationship between features and covariates may overly constrain the representation space.

Inspired by contrastive variational autoencoder [11], we propose a novel *graph disentanglement learning* (GDL) framework to address this challenge. GDL takes graphs constructed from the fMRI data as input, consists of a GNN encoder and 3 following components: an *individual head*, a *covariate head*, and a *disease head*. The features from the disease head are trained to separate the patients from the healthy controls while the covariate head is trained to reflect the covariate information. Different from prior works, we consider all covariates and do not require a linear relationship between features and covariates. The *individual head* is introduced to account for the variability across individuals that cannot be explained by disease or covariates, considering that even healthy controls with the same age and gender will have quite different images. The features from all the 3 heads are concatenated and fed into a reconstruction head to enhance feature representation. Our model was evaluated on 3 datasets — ADHD-200,

SCZ, and Presbycusis — and demonstrated promising classification performance.

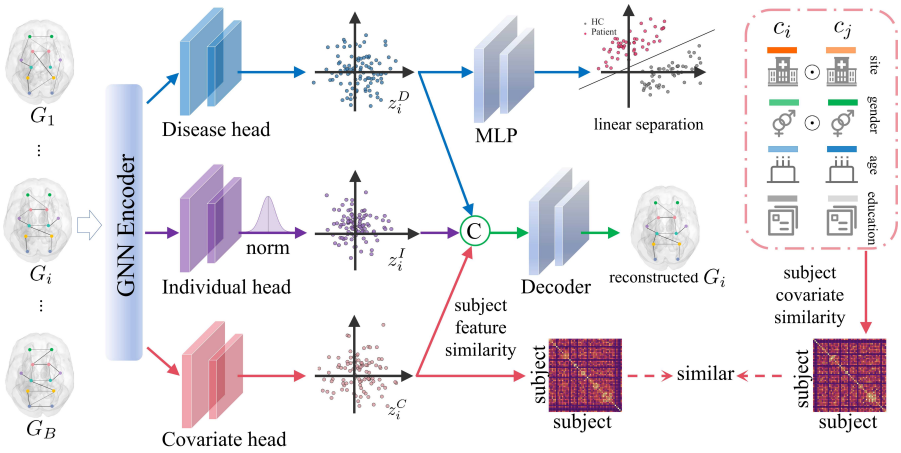

Fig. 1: Overview of the GDL architecture. It is a disentangled representation learning framework comprising 3 components, an individual head, a covariate head, and a disease head, in an encoder-decoder architecture. The disease head is followed by a classification loss to learn disease-related features. The covariate head is followed by a computation of subject feature similarity matrix that is enforced to be similar to the subject similarity matrix from the true covariates.

2 Graph Disentanglement Learning

An overview of the framework is shown in Fig. 1. Suppose we have triplets of graph, covariate, and class label $\{(G_i, c_i, y_i) : i = 1, \ldots, B\}$ from B subjects, where $G_i = (A_i, X_i)$ is a graph of N nodes constructed from the fMRI image of the ith subject with $A_i \in \mathbb{R}^{N \times N}$ being the adjacency matrix and $X_i \in \mathbb{R}^{N \times m}$ being the node features. GDL aims to learn a GNN encoder on the graph that can isolate disease-related features from covariate-related features and individual variations. To achieve this, first, a GNN encoder Enc extracts features $z_i = Enc(G_i), i = 1, \ldots, B$ from these graphs. Then the features go through 3 projection heads: an individual head g_I, a disease head g_D, and a covariate head g_C. The features after the individual head, $z_i^I = g_I(z_i)$, are normalized such that they are from a standard Gaussian distribution. The features after the disease head, $z_i^D = g_D(z_i)$, are followed by a cross entropy loss using the class labels $\{y_i : i = 1, \ldots, B\}$. The features after the covariate head, $z_i^C = g_C(z_i)$, are computed between any 2 subjects to create a $B \times B$ similarity matrix. This similarity matrix is enforced to be close to the similarity matrix derived from the true covariates $\{c_i : i = 1, \ldots, B\}$. Finally, the features from

the 3 heads are concatenated and fed into a decoder *Dec* to reproduced the node features. The details of each component is given below.

Encoder design and feature separation The GNN encoder maps the input graph $G_i = (A_i, X_i)$ to a latent representation z_i. The encoder consists of 2 layers of message passing process and a global average pooling layer. We use the message passing process from the graph isomorphism networks (GINs) [12]: $H_i^{(l)} = MLP\left(((1+\epsilon)I + A_i)H_i^{(l-1)}\right)$, $l = 1, 2$, where $H_i^{(0)} = X_i$, $H_i^{(l)}$ is the updated node representations at layer l. We use $\epsilon = 0$. The multilayer perceptron (MLP) applies to each node with 2 linear layers and ReLU activation in the middle. After the average pooling layer, G_i leads to a feature vector $z_i \in \mathbb{R}^d$.

We assume that z_i contains 3 kinds of information: disease-related information z_i^D, covariate-related information z_i^C, and individual variation z_i^I. To extract these 3 kinds of information, we employ 3 independent MLPs, each with the same architecture but independent parameters, to map z_i into 3 d-dimensional features: $z_i^D = g_D(z_i)$, $z_i^C = g_C(z_i)$, $z_i^I = g_I(z_i)$. These features then go through different branches such that they contain the corresponding information.

Identification of covariate-related features To make the features z_i^C be related to the the covariates, we calculate the subject similarity matrix from these features and make it close to the similarity matrix from the true covariates. Specifically, for features from any two graphs z_i^C, z_j^C, we compute their cosine similarity: $\hat{s}_{ij} = (z_i^C)^\top z_j^C / \|z_i^C\| \|z_j^C\|$. All these cosine similarities form a matrix $\hat{S} = [\hat{s}_{ij}] \in \mathbb{R}^{B\times B}$. This matrix should be close to the similarity matrix from the true covariates.

Suppose that the true covariates are denoted by $c_i = \{c_{ik} : k \in \mathcal{C}\}, i = 1, \ldots, B$, where $\mathcal{C}$ is the set of covariate names, e.g. $\mathcal{C} = \{$age, gender, site, education$\}$. To calculate the similarity matrix $S = [s_{ij}] \in \mathbb{R}^{B\times B}$ from the covariates of all the subjects, we first define the similarity s_{ij} between subject i and j. There are 2 kinds of covariate variables: continuous variable and categorical variable. For continuous variables (e.g. age, years of education), we apply min-max normalization to normalize them in the range $[0, 1]$, and use one minus their difference as the similarity. For categorical variable (e.g. gender, site), we use 1 as their similarity if the 2 categorical variables are the same and 0 otherwise. Hence, for the kth variable, the similarity is defined by

$$s_k(c_{ik}, c_{jk}) = \begin{cases} 1 - |\tilde{c}_{ik} - \tilde{c}_{jk}|, & \text{if } k \text{ continuous} \\ \mathbb{I}(c_{ik} = c_{jk}), & \text{if } k \text{ categorical} \end{cases}, \quad \tilde{c}_{ik} = \frac{c_{ik} - \min(c_{ik})}{\max(c_{ik}) - \min(c_{ik})},$$

where $\mathbb{I}(\cdot)$ is the indicator function: 1 if the inside condition holds and 0 otherwise. The final similarity s_{ij} between subject i and j is a summation of all the components, i.e., $s_{ij} = \sum_{k\in\mathcal{C}} s_k(c_{ik}, c_{jk})$. Then the similarities from any 2 subjects form the similarity matrix $S = [s_{ij}] \in \mathbb{R}^{B\times B}$.

Finally, with $\hat{S}$ and S defined above, we define the covariate loss using the Frobenius norm of the difference between the 2 matrices: $L_{\text{cov}} = \|S - \hat{S}\|_F^2$.

Identification of disease-related features and individual features To learn disease-related features z_i^D, we introduce class labels $\{y_i : i = 1, \ldots, B\}$ to separate $\{z_i^D : i = 1, \ldots, B\}$ in a supervised loss function. Specifically, we use another 2-layer MLP to map z_i^D to logits which are followed by a cross entropy loss L_{sup} using the class labels.

Apart from the covariate-related features and disease-related features, we consider what remain as individual variations. These individual features z_i^I are assumed to follow a standard Gaussian distribution. This is accomplished by normalizing each channel of z_i^I such that the mean is 0 and the standard deviation is 1. This operation gives the normalized individual features z_i^I.

Finally, the 3 features, z_i^D, z_i^C, z_i^I, are concatenated and fed to a decoder Dec to reproduce the node features. The decoder is implemented as another 2-layer MLP. Denote the reproduced node features by $\hat{X}_i = Dec([z_i^D; z_i^C; z_i^I])$. We use a reconstruction loss to ensure that the reconstructed features $\hat{X}_i$ is similar to the original node features X_i: $L_{\text{rec}} = \sum_{i=1}^{B} \|X_i - \hat{X}_i\|_F^2$.

In summary, our loss function is a linear combination of the supervised cross entropy loss, the covariate loss, and the reconstruction loss:

$$L = L_{\text{sup}} + \lambda_1 L_{\text{cov}} + \lambda_2 L_{\text{rec}}.$$

where λ_1 and λ_2 are regularization hyperparameters. This formulation ensures that the learned features effectively disentangle disease-related information, individual variations, and non-imaging covariates.

3 Experiments and Results

3.1 Experimental Setup

Datasets We used 3 fMRI datasets: one publicly available dataset — ADHD-200 — and two private datasets — SCZ and Presbycusis. The fMRI images were preprocessed using the fMRIPrep pipeline [13], which includes reference image estimation, head motion correction, slice timing correction, and susceptibility distortion correction. After aligning the volumes to the MNI152 space, we regressed out confounders such as framewise displacement, global signals, and mean tissue signals. Quality control resulted in retention of 275 ADHD and 205 healthy control (HC) subjects for ADHD-200, 190 HC and 137 SCZ subjects for SCZ, and 112 HC, 130 normal pure-tone audiometry, and 154 presbycusis subjects for the Presbycusis dataset. We then used the AAL1 atlas [14] to divide the brain into 116 regions, and calculated the mean time series (BOLD signals) for each region. The adjacency matrix was computed using Pearson's correlation coefficient between the mean time series of two regions. The node features were derived through the Fourier transform of the mean time series, capturing the total power of 3 low-frequency bands of the ALFFs (Slow-5: 0.01–0.027 Hz, Slow-4: 0.027–0.073 Hz, and Classical: 0.01–0.08 Hz).

Implementation details Our model is implemented in PyTorch and designed to run on a single GPU. The experiments were accelerated using two servers, one equipped with 8 NVIDIA V100 GPUs and the other with 2 NVIDIA A6000 GPUs. The implementation hyperparameters include a learning rate of 0.001, an embedding dimension $d = 32$, a batch size $B = 32$, and $\lambda_1 = 1$, $\lambda_2 = 0.6$. The MLPs for the feature separation ($g_D/g_C/g_I$), cross entropy loss, and decoder have similar structures. All have 2 linear layers with ReLU activation in the middle, with the input and middle dimensions being 32, 64 respectively. The output dimensions for the feature separation, cross entropy loss, decoder are 32, number of classes, and 3×116 respectively. For all the GNN encoders, we use a threshold of zero to set all negative values in the adjacency matrix to zero [15].

Competing methods We compare our proposed GDL framework with 9 competing methods, including 2 supervised learning (SL) methods, BrainGNN [16] and NEGAT [17], 7 self-supervised learning (SSL) methods, GraphCL [18], JOAO [19], LaGraph [20], AGCL [21], GATE [7], BrainGCL [6], and 1 graph disentangling framework, DMG [22]. For the SL methods, we also incorporated a version that takes the adjacency matrix and node features with covariates regressed out in the preprocessing. For the SSL methods, we use a linear SVM [23] for the subsequent disease classification.

We split each dataset into 60% training, 20% validation, 20% test, used the training set to train the models for 300 epochs, selected the best epoch based on the validation set, and reported the accuracy on the test set. We repeated the above procedure 5 times in cross validation. For the SSL methods, following [21], the model was trained on the entire data, but the SVM classifier was only trained on the training set, and the final score was reported on the test set based on the best validation epoch. We used the accuracy and AUC as evaluation metrics, with OvR-AUC [24] for the 3-class classification (Presbycusis) task.

3.2 Results

Classification performance The results show that the SL methods and SSL methods perform similarly across the 3 datasets, with SSL methods generally outperforming SL methods. For the SL methods, there is no significant difference between the original version and the covariate-regressed version. The proposed GDL framework outperforms all methods in accuracy across the 3 datasets. For AUC, GDL also achieves the best performance on Presbycusis and the second best result on SCZ and ADHD.

Ablation studies and sensitivity analysis We removed the individual head (g_I), the covariate head (g_C), and both of them (reducing to a simple supervised GNN with reconstruction) to show their importance. As shown in Table 1, the version without both performs worse than all the competing methods. Removing g_C leads to a more significant performance drop compared to removing g_I, highlighting the critical role of g_C. Overall, both components contribute significantly to our model's performance.

Table 1: Classification results (mean ± std) on 3 datasets using 5 training/validation/test splits. The best training epoch on the validation set is applied to the test set to evaluate the performance. The best performances are highlighted in **bold**, with the second best underlined. The *italicized* methods handle data with covariates regressed out in the preprocessing.

Type	Methods	ADHD		SCZ		Presbycusis	
		Accuracy	AUC	Accuracy	AUC	Accuracy	OvR-AUC
SL	BrainGNN	62.22 ± 3.87	63.04 ± 4.42	64.86 ± 3.57	64.48 ± 2.73	83.09 ± 4.41	84.02 ± 4.51
	BrainGNN	62.24 ± 3.76	62.10 ± 3.85	64.91 ± 4.25	65.02 ± 4.10	83.17 ± 3.76	83.44 ± 3.49
	NEGAT	62.08 ± 2.94	62.95 ± 3.35	63.33 ± 4.56	64.09 ± 3.91	82.79 ± 3.85	82.95 ± 3.36
	NEGAT	62.19 ± 3.14	62.01 ± 3.49	63.44 ± 4.03	63.38 ± 3.98	82.85 ± 4.60	82.20 ± 3.99
SSL	GraphCL	62.86 ± 4.71	62.21 ± 4.27	65.31 ± 4.72	65.19 ± 4.40	83.45 ± 4.49	82.50 ± 3.92
	JOAO	61.92 ± 3.72	62.12 ± 3.84	64.92 ± 5.11	64.80 ± 4.63	83.15 ± 3.77	82.69 ± 4.43
	LaGraph	63.88 ± 2.68	**63.92 ± 2.63**	65.25 ± 3.83	64.46 ± 4.30	83.85 ± 4.02	83.19 ± 3.46
	AGCL	63.25 ± 4.18	62.63 ± 4.29	67.25 ± 4.18	65.35 ± 3.64	84.15 ± 4.17	83.55 ± 3.86
	GATE	62.25 ± 3.84	62.71 ± 4.03	66.28 ± 3.71	66.05 ± 4.12	83.97 ± 3.47	83.27 ± 2.49
	BrainGCL	62.05 ± 3.59	61.52 ± 3.70	65.27 ± 2.97	64.31 ± 3.59	82.85 ± 2.98	83.06 ± 4.14
GDL	DMG	62.57 ± 2.95	62.08 ± 3.51	67.14 ± 3.95	**68.52 ± 4.41**	84.08 ± 3.19	83.95 ± 4.39
	Ours w/o g_I	63.14 ± 3.26	62.91 ± 3.47	67.29 ± 3.12	66.93 ± 3.95	85.18 ± 2.86	85.09 ± 3.42
	Ours w/o g_C	61.87 ± 3.44	61.20 ± 3.57	66.15 ± 3.46	65.72 ± 3.78	85.14 ± 2.93	84.56 ± 4.20
	Ours w/o both	60.62 ± 4.29	60.24 ± 4.75	64.29 ± 3.52	63.86 ± 3.70	80.79 ± 3.81	80.11 ± 4.23
	Ours	**64.38 ± 3.46**	63.09 ± 4.58	**68.95 ± 4.26**	68.28 ± 4.87	**86.28 ± 3.75**	**86.44 ± 5.12**

We tuned the hyperparameters λ_1 and λ_2 both in the range $\{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. As shown in Fig. 2a, higher values of λ_1 and λ_2 generally yield better performance, suggesting that the covariate loss and reconstruction loss are useful. Especially, when λ_1 is sufficiently high, the classification performance remains stable even when λ_2 changes. In 3, 9, 15 out of 25 settings from the 3 datasets respectively, our framework outperforms the second best method with statistical significance, demonstrating the robustness of these hyperparameters.

We also replaced the GNN encoder with GIN, GCN, GAT, GraphSage on the 3 datasets. As shown in Fig. 2b, GIN performs the best among all the GNN encoders across the 3 datasets. Although some GNN encoders show a slight performance drop, most of them still outperform the second best method, demonstrating the robustness of backbone selection.

We tuned the dimension of the latent features d in the range $\{16, 32, 48, 64, 80, 96, 112, 128\}$. As shown in Fig. 2c, 32 is the optimal choice on all the 3 datasets, though 64 leads to better results on SCZ and Presbycusis.

In terms of time cost, our model is also efficient enough, with the smallest training/inference time among all the methods (Fig. 2d).

Interpretability analysis To assess node importance, we computed the gradient of the logits (before the cross entropy loss) with respect to the adjacency matrix. The gradients from all the patients are averaged to produce a global saliency map. Then we sum up each row of this saliency map to get the importance score of each brain region. These importance scores are sorted and the top 10 important regions are visualized in BrainNet viewer [25]. As shown in Fig. 3, the cerebellum plays a significant role in all 3 diseases. In addition, the thalamus

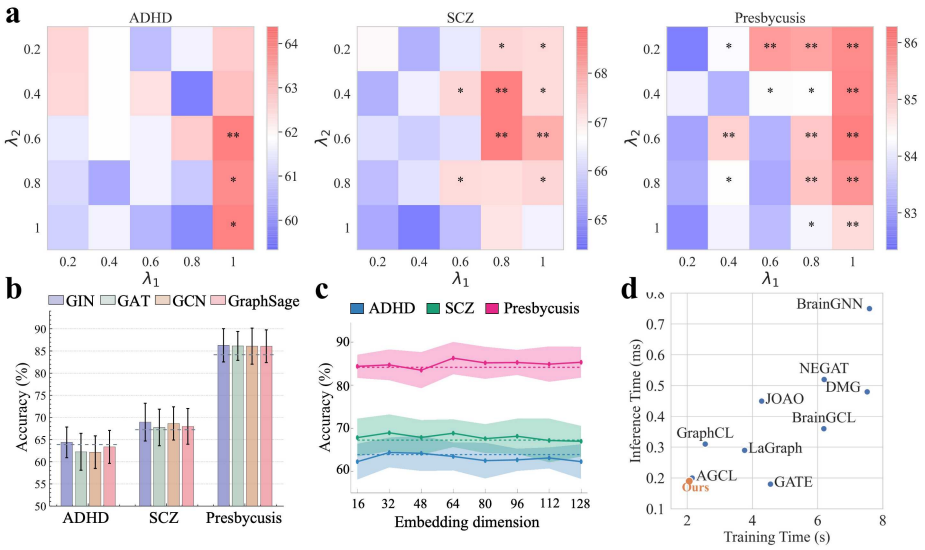

Fig. 2: Ablation studies and sensitivity analysis. (a) The impact of λ_1 and λ_2 in the loss function. */** denotes that the setting is significantly better than the second best method ($p < 0.05/p < 0.01$). (b) The effect of GNN encoder in our framework. The dashed lines represent the best performance achieved by the competing methods. (c) The impact of the embedding dimension of the latent feature. (d) The training time (in second) per epoch and inference time (in millisecond) per instance of all the methods.

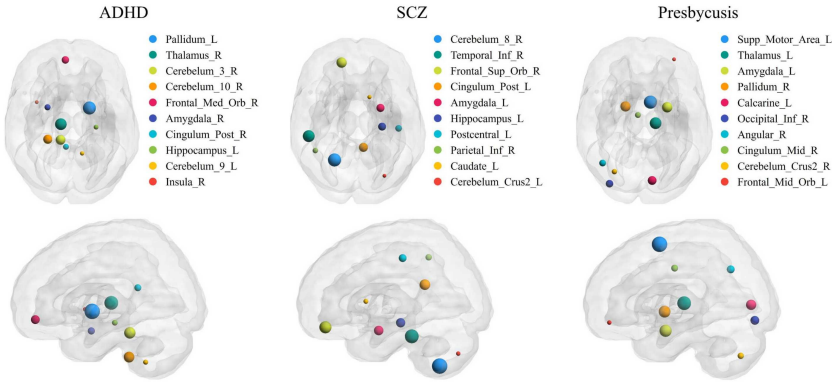

Fig. 3: Top 10 most important brain regions across the 3 datasets. A larger node size indicates that the region has a higher importance score.

and pallidum are implicated in ADHD [26, 27], while the amygdala is associated with SCZ [28, 29] and Presbycusis [30, 31], aligning with clinical findings.

4 Conclusion

In this paper, we proposed a novel graph disentanglement learning (GDL) framework that separates latent features into individual-related, covariate-related, and disease-related components. Our model achieved state-of-the-art classification performance across 3 datasets: ADHD-200, SCZ, and Presbycusis. Ablation studies validated the importance of feature separation and sensitivity analysis showed the robustness of our framework to hyperparameter selection. Interpretability analysis identified disease-related brain regions that align with established clinical findings. These suggest that GDL could be a promising tool in fMRI analysis.

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

Acknowledgments. This work was supported by grants from the National Natural Science Foundation of China (82441016, 82471940), Shanghai Key Laboratory of Child Brain and Development (24dz2260100), and Natural Science Foundation of Shanghai (24TS1415000).

References

[1] Alaa Bessadok, Mohamed Ali Mahjoub, and Islem Rekik. "Graph neural networks in network neuroscience". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.5 (2022), pp. 5833–5848.

[2] KA Smitha et al. "Resting state fMRI: A review on methods in resting state connectivity analysis and resting state networks". In: *The Neuroradiology Journal* 30.4 (2017), pp. 305–317.

[3] Hong Yang et al. "Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI". In: *Neuroimage* 36.1 (2007), pp. 144–152.

[4] Yufeng Zang et al. "Regional homogeneity approach to fMRI data analysis". In: *Neuroimage* 22.1 (2004), pp. 394–400.

[5] Franco Scarselli et al. "The graph neural network model". In: *IEEE Transactions on Neural Networks* 20.1 (2008), pp. 61–80.

[6] Xuexiong Luo et al. "An interpretable brain graph contrastive learning framework for brain disorder analysis". In: *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. 2024, pp. 1074–1077.

[7] Liang Peng et al. "GATE: Graph CCA for temporal self-supervised learning for label-efficient fMRI analysis". In: *IEEE Transactions on Medical Imaging* 42.2 (2022), pp. 391–402.

[8] Ian Goodfellow et al. "Generative adversarial nets". In: *NeurIPS* 27 (2014).

[9] Guangqi Wen et al. "Graph self-supervised learning with application to brain networks analysis". In: *IEEE Journal of Biomedical and Health Informatics* 27.8 (2023), pp. 4154–4165.
[10] Xiaowei Yu et al. "Longitudinal infant functional connectivity prediction via conditional intensive triplet network". In: *MICCAI*. Springer. 2022, pp. 255–264.
[11] Aidas Aglinskas, Joshua K Hartshorne, and Stefano Anzellotti. "Contrastive machine learning reveals the structure of neuroanatomical variation within autism". In: *Science* 376.6597 (2022), pp. 1070–1074.
[12] Keyulu Xu et al. "How powerful are graph neural networks?" In: *arXiv preprint arXiv:1810.00826* (2018).
[13] Oscar Esteban et al. "fMRIPrep: a robust preprocessing pipeline for functional MRI". In: *Nature Methods* 16.1 (2019), pp. 111–116.
[14] Nathalie Tzourio-Mazoyer et al. "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain". In: *Neuroimage* 15.1 (2002), pp. 273–289.
[15] Andrew Zalesky, Alex Fornito, and Ed Bullmore. "On the use of correlation as a measure of network connectivity". In: *Neuroimage* 60.4 (2012), pp. 2096–2106.
[16] Xiaoxiao Li et al. "Braingnn: Interpretable brain graph neural network for fmri analysis". In: *Medical Image Analysis* 74 (2021), p. 102233.
[17] Yuzhong Chen et al. "Adversarial learning based node-edge graph attention networks for autism spectrum disorder identification". In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).
[18] Yuning You et al. "Graph contrastive learning with augmentations". In: *NeurIPS* 33 (2020), pp. 5812–5823.
[19] Yuning You et al. "Graph contrastive learning automated". In: *ICML*. PMLR. 2021, pp. 12121–12132.
[20] Yaochen Xie, Zhao Xu, and Shuiwang Ji. "Self-supervised representation learning via latent graph prediction". In: *ICML*. PMLR. 2022, pp. 24460–24477.
[21] Shengjie Zhang et al. "A-GCL: Adversarial graph contrastive learning for fMRI analysis to diagnose neurodevelopmental disorders". In: *Medical Image Analysis* 90 (2023), p. 102932.
[22] Yujie Mo et al. "Disentangled multiplex graph representation learning". In: *ICML*. PMLR. 2023, pp. 24983–25005.
[23] Rong-En Fan et al. "LIBLINEAR: A library for large linear classification". In: *The Journal of Machine Learning Research* 9 (2008), pp. 1871–1874.
[24] David J Hand and Robert J Till. "A simple generalisation of the area under the ROC curve for multiple class classification problems". In: *Machine learning* 45 (2001), pp. 171–186.
[25] Mingrui Xia, Jinhui Wang, and Yong He. "BrainNet Viewer: a network visualization tool for human brain connectomics". In: *PloS One* 8.7 (2013), e68910.

[26] Larry J Seidman, Eve M Valera, and Nikos Makris. "Structural brain imaging of attention-deficit/hyperactivity disorder". In: *Biological Psychiatry* 57.11 (2005), pp. 1263–1272.

[27] Kerstin Konrad and Simon B Eickhoff. "Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder". In: *Human Brain Mapping* 31.6 (2010), pp. 904–916.

[28] Jem Riffkin et al. "A manual and automated MRI study of anterior cingulate and orbito-frontal cortices, and caudate nucleus in obsessive-compulsive disorder: comparison with healthy controls and patients with schizophrenia". In: *Psychiatry Research: Neuroimaging* 138.2 (2005), pp. 99–113.

[29] Giulio Pergola et al. "The role of the thalamus in schizophrenia from a neuroimaging perspective". In: *Neuroscience & Biobehavioral Reviews* 54 (2015), pp. 57–75.

[30] Chama Belkhiria et al. "Cingulate cortex atrophy is associated with hearing loss in presbycusis with cochlear amplifier dysfunction". In: *Frontiers in Aging Neuroscience* 11 (2019), p. 97.

[31] Jonathan E Peelle et al. "Hearing loss in older adults affects neural systems supporting speech comprehension". In: *Journal of neuroscience* 31.35 (2011), pp. 12638–12643.