MGDR: Multi-Modal Graph Disentangled Representation for Brain Disease Prediction

Bo Jiang[12], Yapeng Li[1], Xixi Wan[1], Yuan Chen[3], Zhengzheng Tu[1(✉)], Yumiao Zhao[1], and Jin Tang[12]

[1] School of Computer Science and Technology, Anhui University, Hefei, China
[2] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
[3] School of Internet, Anhui University, Hefei, China
jiangbo@ahu.edu.cn, zhengzhengahu@163.com

Abstract. In the task of disease prediction, medical data with different modalities can provide much complementary information for disease diagnosis. However, existing multi-modal learning methods often tend to focus on learning shared representation across modalities for disease diagnosis, without fully exploiting the complementary information from multiple modalities. To overcome this limitation, in this paper, we propose a novel Multi-modal Graph Disentangled Representation (MGDR) approach for brain disease prediction problem. Specifically, we first construct a specific modality graph for each modality data and employ Graph Convolutional Network (GCN) to learn node representations. Then, we learn the common information across different modalities and private information of each modality by developing a disentangled representation of modalities model. Moreover, to remove the possible noise from the private information, we employ a contrastive learning module to learn more compact representation of private information for each modality. Also, a new Multi-modal Perception Attention (MPA) module is employed to integrate feature representations of multiple private information. Finally, we integrate both common and private information together for disease prediction. Experiments on both ABIDE and TADPOLE datasets demonstrate that our MGDR method achieves the best performance when compared with some recent advanced methods.

Keywords: Brain disease prediction · Graph learning · Disentangled representation learning · Multi-modal learning.

1 Introduction

Recently,multi-modal brain disease prediction has become a promising approach for Alzheimer's Disease (AD) [1] and Autism Spectrum Disorder (ASD) [2] prediction. Multiple brain imaging modalities such as Magnetic Resonance Imaging (MRI) [3] and Positron Emission Tomography (PET) [4] can provide complementary structural and functional information for abnormal brain regions [3, 5]. In addition, various clinical data such as cognitive tests, demographic information, etc, are also helpful to predict brain diseases [6]. By integrating multi-modal

data [7, 8], we can gain a more holistic view of brain function and structure, enhancing the understanding of brain disorders.

Also, deep learning methods have made significant progress in medical field [9, 10]. In particular, Graph Neural Networks (GNNs) are extensively utilized in the field of biomedical analysis [11] including AD and ASD [6] [12] prediction. For example, Kazi et al. [13] combine imaging data and non-imaging data to construct an overall affinity graph of subjects for the disease prediction tasks. Pariost et al. [12] utilize non-imaging data such as age and gender with imaging information to construct a relationship graph for patients to enhance the learning performance. Song et al. [10] propose a multi-center attention graph to integrate multi-modal data together for composition and subsequently employ multiple-channel pooling GCN for disease prediction. In addition to single graph model, some researchers have also conducted many studies based on multi-graph models [14–16]. Compared with single graph model, multiple graphs can effectively exploit the feature relationships for multi-modality data. For example, Kazi et al. [15] propose LSTM-GCN method to construct multiple graphs from multi-modal data, treating these graphs as a sequence and using attention mechanism to integrate multi-modal features for final decision. Wen et al. [16] propose to construct multiple graphs for brain networks to obtain richer brain structure information for ASD diagnosis. Although existing GNN-based methods have been widely utilized for disease prediction, they generally fail to capture the intrinsic relationships of different modalities [6]. In other words, they generally exploit the common information of modalities, failing to fully consider the complementary information in the private information of each modality.

To address these issues, we propose a novel Multi-modal Graph Disentangled Representation (MGDR) approach for brain disease prediction. To be specific, MGDR first represents multi-modal information of the subjects as multiple graphs and utilizes Graph Convolution Network (GCN) to extract features of subjects. Then, inspired by works [17, 18], MGDR disentangles multi-modal graphs to obtain the common information between modalities and private information of each modality respectively. Particularly, to extract reliable information, we employ Singular Value Decomposition (SVD) to decouple the common information. Also, to filter out noises from the private information, we utilize a contrastive learning module to constrain the representations of private information. Finally, MGDR integrates both common and private information to predict brain disease. The main contributions are summarized as follows:

- We propose to develop a novel multi-graph representation learning approach that fully exploits the dependencies of subjects to learn context-aware feature representation for each subject.
- We propose to exploit a multi-modal disentangled representation pipeline by considering both common information of modalities and private information of each modality to better guide disease prediction.
- Comparing with recent methods, our proposed MGDR method achieves the best prediction performance on the public ABIDE and TADPOLE datasets.

2 Related Work

For brain disease prediction with multi-modal data, existing researches [7,12,13, 17] have achieved significant progress. For example, graph-based learning methods for predicting brain diseases have been widely applied [6, 14, 19, 20]. Among them, Cosmo et al. [20] propose an end-to-end graph learning framework to learn the optimal single graph. Kazi et al. [14] propose to construct multiple affinity graphs for sharing a node set to represent the subjects. Zheng et al. [6] propose the MMGL, which utilizes multi-modal attention to fuse features from different modalities and then constructs a global graph for downstream tasks. In addition to graph-based learning methods described above, several other approaches have been proposed to predict brain diseases [7,17,21]. For example, Xu et al. [7] propose a method called MSTGC, which utilizes two contrastive constraints to mine the complementary information between multiple modalities and the unique information of each modality. Wang et al. [17] propose a method DMAAN to achieve latent representations of different modalities via encoding multi-modal imaging data. Especially, the DMAAN [17] decomposes multi-modal data into the shared and the specific representations to predict brain disease, being considered more related work with our proposed MGDR. Although the above methods obtain relatively good performance for brain disease prediction, they cannot effectively exploit the dependences between different modalities. There is still a lack of exploration for complementary information between multiple modalities, which hinders the full utilization of the shared information of modalities and the private information of each modality.

3 Methodology

The overall framework of the proposed MGDR is shown as Fig. 1. Firstly, we utilize multi-graph representation learning to obtain the feature representation of each subject across multiple modalities. Then, we disentangle multi-modality features to obtain rich common information between modalities and private information of each modality. Finally, we integrate both common and private information together for disease prediction.

3.1 Multi-Graph Representation Learning

Let $\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)} \cdots \mathbf{X}^{(M)}\}$ denotes the original features of different modalities where M is the modality number and $\mathbf{X}^{(m)} = \{x_1^{(m)} \cdots x_N^{(m)}\} \in \mathbb{R}^{N \times d_m}$ denotes the subjects' features of the m-th modality. To capture the relationships between subjects in multiple modalities, we construct multiple graphs. Specifically, we utilize cosine similarity [6] to calculate the relationships between subjects in each modality as follows:

$$\mathbf{A}_{ij}^{(m)} = \mathrm{sim}\big(x_i^{(m)}, x_j^{(m)}\big) = \cos\big(x_i^{(m)}\mathbf{W}^{(m)}, x_j^{(m)}\mathbf{W}^{(m)}\big), \tag{1}$$

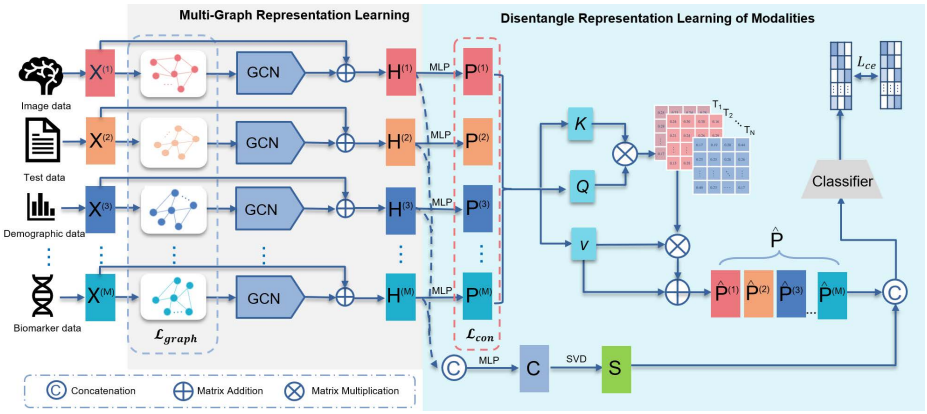


Fig. 1. The DMGR framework mainly includes Multi-Graph Representation Learning (MGRL) and Disentangle Representation of Modalities (DRM).

where $\mathbf{W}^{(m)}$ denotes the transformation matrix. To encourage the smoothness and sparsity of the learned graph, we add regularization constraint $\mathcal{L}_{graph}$ on $\mathbf{A}$. Here, the Dirichlet energy function [22] is utilized to optimize the learned graphs and some additional regularization terms [6] are added to avoid the trivial solution and excessive sparsity, i.e.,

$$\mathcal{L}_{graph}(\mathbf{A}) = \tag{2}$$
$$\frac{1}{M}\sum_{m=1}^{M}\Big(\frac{1}{2N^2}\sum_{i,j=1}^{N}\mathbf{A}_{ij}^{(m)}\|x_i^{(m)}-x_j^{(m)}\|^2+\big(-\frac{\alpha}{N}\mathbf{1}^{\mathbf{T}}\log(\mathbf{A}^{(m)}\mathbf{1})\big)+\frac{\beta}{N^2}\|\mathbf{A}^{(m)}\|_F^2\Big)$$

where $\mathbf{1}$ represents a vector of all one and $\|\cdot\|$ denotes the Frobenius norm. α and β are two hyper-parameters to balance different terms. To better learn the representation of each subject and capture the dependencies of different subjects, we use Graph Convolutional Network (GCN) [23] to obtain context-aware node representations as follows:

$$\mathbf{H}^{(m,l+1)} = \mathrm{ReLU}(\hat{\mathbf{D}}^{(m)-\frac{1}{2}}\hat{\mathbf{A}}^{(m)}\hat{\mathbf{D}}^{(m)-\frac{1}{2}}\mathbf{H}^{(m,l)}\Theta^{(m,l)}), \tag{3}$$

where $\hat{\mathbf{A}}^{(m)} = \mathbf{A}^{(m)} + w\mathbf{I}$ where w indicates the weight coefficient. $\Theta^{(m)}$ denotes the trainable weight matrix and $\hat{\mathbf{D}}$ represents the degree matrix of $\hat{\mathbf{A}}$.

3.2 Disentangled Representation of Modalities

Let $\mathbf{H} = \{\mathbf{H}^{(1)}\cdots\mathbf{H}^{(M)}\}$ denotes the residual weighted sum of the obtained feature representations via above GCN module and the original feature representations. To learn richer common and private information of multiple graphs, we propose to develop a novel Disentangled Representation of Modalities (DRM) module. There are two main steps of the proposed DRM module.

Common Information Representation. To obtain more effective common information, we first concatenate the feature representations of all modalities together and then apply MLP as follows,

$$\mathbf{C} = \mathrm{MLP}\big(\mathrm{Con}(\mathbf{H}^{(1)}, \mathbf{H}^{(2)} \cdots \mathbf{H}^{(M)})\big), \tag{4}$$

where $\mathrm{Con}(\cdot)$ denotes the concatenation operation and $\mathbf{C}$ denotes the extracted common information. Also, we disentangle the fused common information $\mathbf{C}$ by using SVD [18] to obtain its left singular vectors $\mathbf{U}$ and right singular vectors $\mathbf{V}$. By retaining the first k left singular vectors and the first k right singular vectors, the common information $\mathbf{S}$ is obtained via $\mathbf{S} = \mathbf{U}_k \mathbf{V}_k^T$. This approach is beneficial as it effectively captures the principal directions of variation in the data, while preserving the most crucial structural information and reducing the impact of noise.

Private Information Representation. Private information contains the complementary information of different modalities and also probably contains some noises [18]. To promote the disentanglement of private information, we employ a MLP module with non-shared parameters. This operation can enhance the feature representations $\mathbf{H}^{(m)}$ to obtain the private information $\mathbf{P}^{(m)}$ as

$$\mathbf{P}^{(m)} = \mathrm{MLP}(\mathbf{H}^{(m)}). \tag{5}$$

To further reduce the possible noises, we also employ a contrastive learning module [18]. Specifically, Let $(\mathrm{p}_i^{(m)}, \mathrm{p}_j^{(m)}) \in \mathcal{E}_{pos}^{(m)}$ denote the pair with high similarity which is considered as the positive pair, while $(\mathrm{p}_k^{(m)}, \mathrm{p}_l^{(m)}) \in \mathcal{E}_{neg}^{(m)}$ denote the subject pair with low similarity which is considered as the negative pair. Thus, the contrastive loss is designed as follows:

$$\mathcal{L}_{con} = -\sum_{m=1}^{\mathrm{M}} \log \frac{\sum e^{\theta(\mathrm{p}_i^{(m)}, \mathrm{p}_j^{(m)})/\tau}}{\sum e^{\theta(\mathrm{p}_i^{(m)}, \mathrm{p}_j^{(m)})/\tau} + \sum e^{\theta(\mathrm{p}_k^{(m)}, \mathrm{p}_l^{(m)})/\tau}}, \tag{6}$$

where θ denotes the metric function and τ represents the temperature parameter. Besides, inspired by work [6, 24], to integrate multiple private information, we introduce Multi-modal Perception Attention (MPA) module to balance the importance of different modalities. To be specific, we first utilize three linear transformations to project $\mathbf{P}^{(m)}$ into queries $\mathbf{Q}^{(m)} = \{q_1^{(m)} \cdots q_N^{(m)}\}$, keys $\mathbf{K}^{(m)} = \{k_1^{(m)} \cdots k_N^{(m)}\}$ and values $\mathbf{V}^{(m)} = \{v_1^{(m)} \cdots v_N^{(m)}\}$. Then, we calculate the Multi-modal Perception Attention weights between different modalities as:

$$\mathbf{T}_j^{(m,n)} = \frac{\exp\left((q_j^{(m)})^T k_j^{(n)} / \sqrt{d_f}\right)}{\sum_{n=1}^{M} \exp\left((q_j^{(m)})^T k_j^{(n)} / \sqrt{d_f}\right)}. \tag{7}$$

To aggregate the information from other modalities and enhance the feature representation of each modality, we conduct the message propagation as,

$$\hat{\mathrm{p}}_j^{(m)} = \sum\nolimits_{n=1}^{M} \mathbf{T}_j^{(m,n)} v_j^{(n)} + \gamma v_j^{(n)}, \tag{8}$$

where $\gamma > 0$ denotes the weight parameter.

Feature Fusion. We concatenate features $\{\hat{\mathrm{p}}_1^{(m)}, \cdots \hat{\mathrm{p}}_N^{(m)}\}$ to obtain the unified feature representations $\hat{\mathbf{P}}^{(m)}$ of all subjects in the m-th modality. The disentangled common information $\mathbf{S}$ and private information $\hat{\mathbf{P}} = \mathrm{Con}\big(\hat{\mathbf{P}}^{(1)} \cdots \hat{\mathbf{P}}^{(M)}\big)$ are fused as follows:

$$\mathbf{Z} = \mathrm{Con}\big(\hat{\mathbf{P}}, \mathbf{S}\big). \tag{9}$$

We feed the fused features $\mathbf{Z}$ into the classifier to obtain predicted labels $\hat{Y} = (\hat{y}_1, \hat{y}_2 \cdots \hat{y}_N)$ for all subjects. The proposed method is training in an end-to-end manner. The total training loss for our MGDR model can be defined as:

$$\mathcal{L}_{total} = \delta \mathcal{L}_{graph}(\mathbf{A}) + \zeta \mathcal{L}_{con} + \eta \mathcal{L}_{ce}(Y, \hat{Y}), \tag{10}$$

where $\mathcal{L}_{ce}(Y, \hat{Y})$ represents the cross-entropy loss and δ, ζ, η are parameters.

4 Experiments and Results

4.1 Datasets and Implementation Details

ABIDE: The Autism Brain Imaging Data Exchange (ABIDE) dataset [25] is a public dataset. After pre-processing, we obtain neuroimaging and phenotypic data from subjects with ASD, including four modalities, e.g., fMRI connection networks, automated anatomical quality assessment, automated functional quality assessment and demographic information. **TADPOLE:** The TADPOLE dataset [26] comes from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Six modalities are acquired through data preprocessing which include cerebrospinal fluid (CSF) biomarkers, PET, cognitive tests, demographic information, MRI and risk factors. The demographic information is shown in Table 1 on both ABIDE and TADPOLE dataset.

Table 1. Demographic information of subjects on both ABIDE and TADPOLE datasets. MOCA and MMSE are two cognitive function screening scales.

TADPOLE	Age	Female/Male	MMSE	MoCA
AD	73.29±7.97	30/44	22.82±2.93	16.86±5.06
MCI	70.87±7.19	144/171	28.14±1.70	23.53±3.10
NC	72.81±5.96	114/95	29.13±1.11	25.93±2.45

ABIDE	Age	Female/Male	Open-Eye	Closed-Eye
ASD	17.07±7.95	54/349	288	115
NC	16.84±7.23	90/378	321	147

Data Preprocessing and Implementation Details. Collecting complete multi-modal features for brain disease prediction is a challenging task. On the

Table 2. Comparisons of our MGDR method with state-of-the-art methods.

	TADPOLE		ABIDE			
METHOD	ACC(%)	AUC(%)	ACC(%)	AUC(%)	SEN(%)	SPE(%)
InceptionGCN [13]	77.42±1.53	81.58±1.31	72.69±2.37	72.81±1.94	80.29±5.10	74.41±6.22
MLP	82.28±4.39	83.13±3.20	75.22±8.06	79.30±7.95	77.35±9.00	75.24±10.9
PopGCN [12]	82.37±5.10	80.71±4.21	69.80±3.35	70.32±3.90	73.35±7.74	80.27±6.48
LSTMGCN [15]	83.40±4.11	82.42±7.97	74.92±7.74	74.71±7.92	78.57±11.6	78.87±7.79
Multi-GCN [14]	83.50±4.91	89.34±5.38	69.24±5.90	70.04±4.22	70.93±4.68	74.33±6.07
EV-GCN [19]	88.51±2.34	89.97±2.15	85.90±4.47	84.72±4.27	88.23±7.18	79.90±7.37
LGL [20]	91.37±2.12	93.96±1.45	86.40±1.63	85.88±1.75	86.31±4.52	88.42±3.04
MMGL [6]	92.31±1.73	93.91±2.10	89.77±2.72	89.81±2.56	**90.32±4.21**	89.30±6.04
MAFGN [8]	92.80±0.92	93.32±2.10	-	-	-	-
MGDR(Ours)	**93.64±3.90**	**94.89±2.96**	**91.39±2.00**	**91.25±2.07**	89.33±4.55	**93.16±3.27**

ABIDE dataset, we preprocess the multi-modality data using the Preprocessed Connectome Project (PCP) [12, 27, 28]. On the TADPOLE dataset, we adopt the approach [6] for feature selection, which takes advantage of morphological features. Then, we compute the missing rate of features for each subject. Features of each subject with a missing rate above 5% are excluded. For subjects with a missing rate below 5%, we address the missing values by imputing the mean of the available data. Our experiment utilizes a 10-fold cross-validation method which can avoid overfitting and obtain more reliable results.

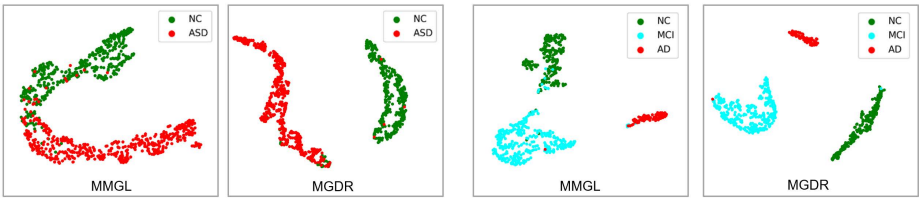
4.2 Comparison Results

Quantitative Results. We utilize Accuracy (ACC), Area Under the Curve (AUC), Specificity (SPE) and Sensitivity (SEN) for evaluation. We compare the proposed method with nine recent methods: Multi-Layer Perceptron (MLP), single-graph-based methods including PopGCN [12], InceptionGCN [13], EV-GCN [19], LGL [20], MMGL [6], MAFGN [8], and multi-graph-based methods including Multi-GCN [14], LSTMGCN [15]. The comparison results are summazired in Table 2. Here, we can observe that our MGDR model achieves better performance than existing single-graph-based methods and multi-graph-based methods. Specifically, on the TADPOLE dataset, our MGDR method outperforms other methods, with improvements of 0.84% and 0.93% over the second-best method on ACC and AUC, respectively. On the ABIDE dataset, compared with the second-best method, our method shows improvements of 1.62%, 1.44% and 3.86% on ACC, AUC and SPE, respectively. However, our method exhibits a relatively high standard deviation because the utilized SVD may result in some fluctuations in the model's performance.

Ablation Study. We perform ablation experiments on both ABIDE and TADPOLE datasets. The results are shown in Table 3. Comparison experiments are designed as follows: (1) Backbone: We only retain MPA [6] as our backbone. The initial features are input into the MPA module. Then, the results are predicted after modal interaction. (2) Backbone+$\hat{\mathbf{P}}$: We utilize MGRL and MPA module to obtain private information $\hat{\mathbf{P}}$ of each modality. The results of this model have been significantly improved, indicating that the learned private

Table 3. Results of the ablation study on ABIDE and TADPOLE datasets.

	TADPOLE		ABIDE			
METHOD	ACC(%)	AUC(%)	ACC(%)	AUC(%)	SEN(%)	SPE(%)
Backbone	89.12±3.31	90.84±2.63	86.48±2.99	86.40±3.30	85.34±3.57	87.45±3.14
Backbone+$\hat{\mathbf{P}}$	92.63±2.65	92.80±2.26	91.05±1.90	90.92±1.96	89.10±4.39	92.73±3.17
Backbone+$\hat{\mathbf{P}}$+$\mathcal{L}_{con}$	92.80±3.21	93.53±3.31	91.16±1.62	91.14±1.67	**89.57±3.94**	92.51±2.92
MGDR(Backbone+$\hat{\mathbf{P}}$+$\mathcal{L}_{con}$+$\mathbf{S}$)	**93.64±3.90**	**94.89±2.96**	**91.39±2.00**	**91.25±2.07**	89.33±4.55	**93.16±3.27**



(a) ABIDE (b) TADPOLE

Fig. 2. Visualization of the feature representations on different datasets.

information $\hat{\mathbf{P}}$ is effective. (3) Backbone+$\hat{\mathbf{P}}$+$\mathcal{L}_{con}$: Based on the Backbone+$\hat{\mathbf{P}}$, incorporating the contrastive constraint $\mathcal{L}_{con}$ in Eq.(6) can obtain a good performance. This experiment demonstrates the effectiveness of reducing noise within the graphs by $\mathcal{L}_{con}$. (4) Backbone+$\hat{\mathbf{P}}$+$\mathbf{S}$+$\mathcal{L}_{con}$: This is our MGDR model. After extracting the private information $\hat{\mathbf{P}}$, we incorporate the disentangled common information $\mathbf{S}$. Our MGDR model shows the best performance, indicating the effectiveness of integrating the disentangled common and private information.

Visualization. To evaluate the performance of our MGDR method on both ABIDE and TADPOLE dataset, we use the 2D t-SNE [29] to visualize the feature representation in MMGL [6] and in our method, separately. As shown in Fig.2 (a) and (b), MGDR method obtains superior performance and exhibits smaller intra-class distances and larger inter-class distances, which indicate that our method can obtain better feature representations in the subsequent tasks.

5 Conclusion

This paper proposes a novel Multi-modal Graph Disentangled Representation (MGDR) method for brain disease prediction. MGDR has two main aspects. It first employs Multi-Graph Representation Learning (MGRL) module to capture the dependences of multi-modal subjects for context-aware subject representation. Subsequently, the Disentangled Representation of Modalities (DRM) is employed to extract common and private information respectively for brain disease diagnosis. The experiments demonstrate that the proposed method outperforms other state-of-the-art methods on public ABIDE and TADPOLE datasets.

Acknowledgments. This work was supported by National Natural Science Foundation of China (62076004, 62376005); the University Synergy Innovation Program of Anhui Province under Grant GXXT-2022-032 and GXXT-2022-014.

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

References

1. Zhang, D., Wang, Y., Zhou, L., et al.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage, **55**(3), 856-867 (2011)
2. Arbabshirani, M. R., Plis, et al.: Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage, **145**, 137-165 (2017)
3. Hu, D., Zhang, H., Wu, Z., Wang, F., Wang, L., Smith, J. K., et al.: Disentangled-multimodal adversarial autoencoder: Application to infant age prediction with incomplete multimodal neuroimages. IEEE Transactions on Medical Imaging, **39**(12), 4137-4149 (2020)
4. JXue, Z., Li, P., Zhang, L., Lu, X., Zhu, G., Shen, P., et al.: Multi-modal co-learning for liver lesion segmentation on PET-CT images. IEEE Transactions on Medical Imaging, **40**(12), 3531-3542 (2021)
5. Yang, Y., Ye, C., Guo, X., Wu, T., Xiang, Y., Ma, T.: Mapping Multi-Modal Brain Connectome for Brain Disorder Diagnosis via Cross-Modal Mutual Learning. IEEE Transactions on Medical Imaging, **43**, 108-121 (2023)
6. Zheng, S., Zhu, Z., et al.: Multi-Modal Graph Learning for Disease Prediction. IEEE Transactions on Medical Imaging, **41**, 2207-2216 (2021)
7. Xu, R., Zhu, Q., Li, S., Hou, Z., Shao, W., Zhang, D.: MSTGC: Multi-Channel Spatio-Temporal Graph Convolution Network for Multi-Modal Brain Networks Fusion. IEEE Transactions on Neural Systems and Rehabilitation Engineering, **31**, 2359-2369 (2023)
8. Yang, F., Wang, H., Wei, S., Sun, G., Chen, Y., Tao, L.: Multi-model adaptive fusion-based graph network for Alzheimer's disease prediction. Computers in Biology and Medicine, **153**, 106518-106528 (2023)
9. Li, H., Shi, X., Zhu, X., Wang, S., Zhang, Z.: FSNet: Dual Interpretable Graph Convolutional Network for Alzheimer's Disease Analysis. IEEE Transactions on Emerging Topics in Computational Intelligence, **7**(1), 15-25 (2022)
10. Song, X., Zhou, F., et al.: Multicenter and Multichannel Pooling GCN for Early AD Diagnosis Based on Dual-Modality Fused Brain Network. IEEE Transactions on Medical Imaging, **42**(2), 354-367 (2022)
11. Al-Sabri, R., Gao, J., et al.: Multi-view graph neural architecture search for biomedical entity and relation extraction. IEEE/ACM Transactions on Computational Biology and Bioinformatics, **20**(2), 1221-1233 (2022)
12. Parisot, S. et al.: Spectral Graph Convolutions for Population-Based Disease Prediction. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D., Duchesne, S. (eds.) MICCAI 2017, LNCS, vol. 10435, pp. 177-185. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_21
13. Kazi, A. et al.: InceptionGCN: Receptive Field Aware Graph Convolutional Network for Disease Prediction. In: Chung, A., Gee, J., Yushkevich, P., Bao, S. (eds.) IPMI 2019, LNCS, vol. 11492. pp. 73-85. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20351-1_6
14. Kazi, A., Shekarforoush, S., Kortuem, K., Albarqouni, S., Navab, N.: Self-attention equipped graph convolutions for disease prediction. In: International Symposium on Biomedical Imaging, pp. 1896-1899, IEEE (2019). https://doi.org/10.1109/ISBI.2019.8759274

15. Kazi, A. et al.: Graph Convolution Based Attention Model for Personalized Disease Prediction. In: Shen, D., et al. (eds.) MICCAI 2019, LNCS, vol. 11767, pp. 122-130. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_14
16. Wen, G., Cao, P., Bao, H., Yang, W., Zheng, T., Zaiane, O.: MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. Computers in Biology and Medicine, **142**, 105239-105253 (2022)
17. Wang, T., Chen, X., Zhang, J., Feng, Q., Huang, M.: Deep multimodality-disentangled association analysis network for imaging genetics in neurodegenerative diseases. Medical Image Analysis, **88**, 102842-102858 (2023)
18. Mo, Y., Lei, Y., Shen, J., Shi, X., Shen, H.T., Zhu, X.: Disentangled Multiplex Graph Representation Learning. In: International Conference on Machine Learning 2023, pp. 24983-25005 (2023)
19. Huang, Y., Chung, A. C.: Edge-variational graph convolutional networks for uncertainty-aware disease prediction. In: Martel, A.L., et al. (eds.) MICCAI 2020, LNCS, vol. 12267, pp. 562-572. Springer, Cham (2020) https://doi.org/10.1007/978-3-030-59728-3_55
20. Cosmo, L., Kazi, A., Ahmadi, SA., Navab, N., Bronstein, M.: Latent-Graph Learning for Disease Prediction. In: Martel, A.L., et al. (eds.) MICCAI 2020, LNCS, vol. 12262, pp. 643-653. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_62
21. Zhang, J., Zhou, L., Wang, L., Liu, M., Shen, D.: Diffusion kernel attention network for brain disorder classification. IEEE Transactions on Medical Imaging, **41**(10), 2814-2827 (2022)
22. Belkin, Mikhail, and P. Niyogi.: Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In: Advances in Neural Information Processing Systems (2001), pp.585-591 (2001)
23. Kipf, T., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks. In: International Conference on Learning Representations (2017)
24. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser,Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
25. Di Martino, A., Yan, C. G., Li, Q., Denio, E., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry, **19**(6), 659-667 (2014)
26. Marinescu, R.V., Oxtoby, N.P., Young, A.L., Bron, E.E., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Klein, S., Alexander, D.C., Consortium, T.E., Initiative, F.T.: TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease. arXiv preprint arXiv:1805.03909 (2018)
27. Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., Khundrakpam, B.S., Lewis, J.D., Li, Q., Milham, M., Yan, C., Bellec, P.: The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. In: Neuroinformatics (2013)
28. Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., Frackowiak, R. S.: Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping, **2**(4), 189-210 (1994)
29. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. Journal of Machine Learning Research, **9**(11), 2579–2605 (2008)