This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

# SPromptGL: Semantic Prompt Guided Graph Learning for Multi-modal Brain Disease

Xixi Wan[1 [0009−0006−6254−6713]], Bo Jiang[2,3 (✉) [0000−0002−6238−1596]], Shihao Li[1 [0009−0001−0923−3965]], and Aihua Zheng[1 (✉) [0000−0002−9820−4743]]

[1] School of Artificial Intelligence, Anhui University, Hefei, China
[2] School of Computer Science and Technology, Anhui University, Hefei, China
[3] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
zeyiabc@163.com, ahzheng214@foxmail.com

**Abstract.** Multi-modal brain disease diagnosis provides a more robust and comprehensive prediction of diverse diseases by integrating medical data from different modalities. However, recent methods generally fail to account for the modality-specific discriminant regions in semantic information, which causes models to focus on non-lesion areas while neglecting the actual lesion regions. To address this issue, we propose Semantic Prompt-guided Graph Learning (SPromptGL), a novel approach for multi-modal disease prediction that captures the discriminative regions of different modalities while enhancing their interaction and fusion. Firstly, to explore the relationship between subjects of different modalities, we propose constructing an interactively multi-relation graph for multi-modal data. It is dynamically learned by designing graph learning loss terms. The multi-layer graph convolutional neural network is utilized to learn context-enriched representations for each subject. Then, to better capture the significant region representations of different modalities, we propose a semantic prompt-guided learning network to excavate the modality-specific lesion regions of related diseases. Specifically, a set of semantic prompts of related brain diseases is first guided to capture fine-grained local details to enhance patch representation. And then we couple with a relation-aware embedding strategy to refine discriminative features. Compared with state-of-the-art methods, our approach achieves superior performance on different benchmark datasets. Code is available at https://github.com/wanxixi11/SPromptGL.

**Keywords:** Brain Disease Prediction · Multi-relation Graph · Graph Convolutional Neural · Prompt-guided Learning Network.

## 1 Introduction

The treatment and prevention of Alzheimer's Disease (AD) [3, 32] and Autism Spectrum Disorder (ASD) [1, 10] have gained significant attention, with notable advancements in computer-aided prediction using deep learning methods, such as Graph Neural Networks (GNNs) [15, 16, 32], Convolutional Neural Networks (CNNs) [17, 28] and Recurrent Neural Networks (RNNs) [25, 31]. Among that,

Kazi *et al.* [13] propose InceptionGCN that fuses multi-modal data to build an overall graph for the disease prediction problem. Zheng *et al.* [32] also propose to predict brain diseases by constructing a global graph with multi-modal features. Zhou *et al.* [33] propose the GIGCN method, which separately constructs two graphs by capturing dual relationships among regions of interest in the brain. Although the aforementioned graph-based methods for brain disease prediction use various graphs (e.g., region, sample), they fail to effectively capture multi-modal data dependencies and interactions.

On the other hand, Large Language Models (LLMs) [2, 6] have been widely applied in the diagnosis of brain diseases, facilitating the development of precise and robust models while reducing the reliance on large-scale annotated datasets. Unlike natural images, brain disease-related data typically exhibit pathological relevance only in specific localized regions. For example, studies have demonstrated a significant correlation between the hippocampus and AD [24, 33]. In recent years, some approaches have employed region-based prompts to guide foundational models in focusing on critical areas within medical images [6, 24]. However, most existing methods primarily rely on simple local feature weighting, failing to fully explore the fine-grained information of pathological regions within the modality and diverse relationships between local details [6, 24, 27].

To overcome the aforementioned issues, we propose a novel Semantic Prompt-guided Graph Learning (SPromptGL) for multi-modal-based brain disease prediction. The core idea of our SPromptGL is to formulate multi-modal information of the brain as a multi-relation graph representation models relationships of subjects via leveraging multi-modal data and then develop a novel prompt-guided learning network to capture the discriminant regions for the brain disease prediction task. In detail, we first propose a new interactively multi-relation graph strategy to learn a more effective graph with semantic constraints for multi-modal data. Then, a multi-layer graph convolutional neural network is employed to learn context-enhanced feature representation for each subject. Finally, to identify the discriminating lesion area and reduce the noise effect of non-lesion areas, a prompt-guided embedding network is designed to explore the modality-specific fine-grained lesion regions of related diseases. Specifically, a set of semantic prompts of related brain diseases is generated to enhance patch representations, and we couple them with a relation-aware embedding strategy to refine discriminative local contexts. Overall, the main contributions of this paper are summarized as follows:

- We propose to employ a multi-relation graph representation for brain disease prediction tasks, where each relation optimizes multi-modal data. The proposed method can not only utilize features of each modality but also fully exploit the dependencies and interactions of different modalities.
- We propose a prompt-guided embedding network, which first guides semantic cues to obtain discriminant local information and then embeds fine-grained local details obtained into global contexts for better exploiting each modal representation for each subject. This proposed strategy can distinguish the

modality-specific lesion regions of related diseases and reduce the noise interference in the no-lesions, improving brain disease diagnosis.
- Experimental results on the commonly used TADPOLE and ABIDE datasets demonstrate that the proposed method is superior to other state-of-the-art methods for predicting brain diseases.

## 2 The Proposed Method

We propose a semantic prompt-guided learning network with a graph-based method for brain disease prediction, capturing discriminative regions and modeling cross-modal relationships to improve feature representation, as shown in Fig. 1. Firstly, we conduct feature selection on the morphological features of each modality to extract complete multi-modal features for each subject. Subsequently, A multi-relation graph representation is introduced to model diverse subject connections, followed by a multi-layer graph convolutional network to learn context-enriched feature representations. Finally, we propose a prompt-guided embedding network to enhance modality-specific representations. This strategy leverages semantic prompts related to brain diseases to capture discriminative local features, which are then integrated into global contexts by modeling their relationships to refine features of multi-modal data.

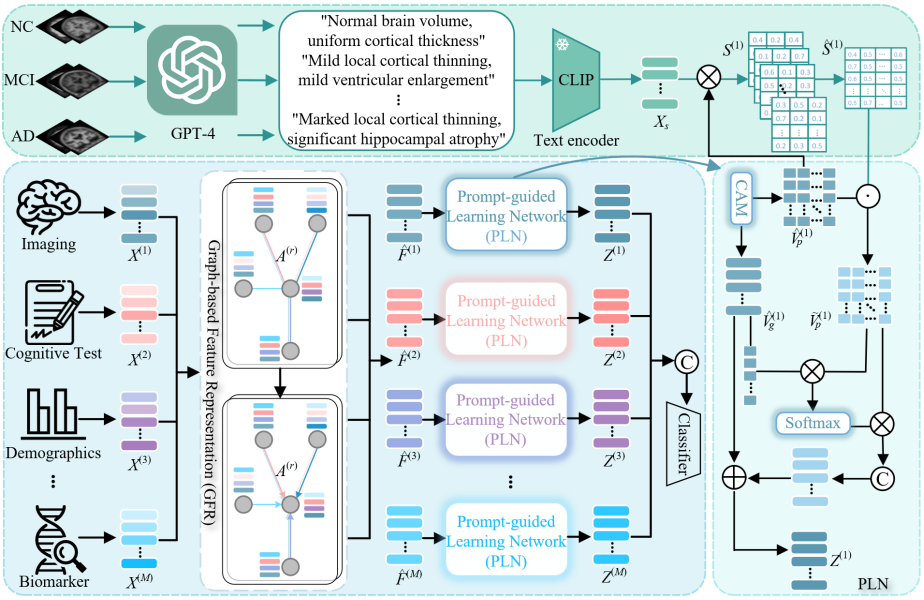


**Fig. 1.** The overall framework of our proposed method, which includes Graph-based Feature Representation (GFR) and Prompt-guided Learning Network (PLN).

### 2.1 Graph-based Feature Representation

Let $X = \{X^{(1)}, X^{(2)} \cdots X^{(M)}\}$ denote the obtained relatively complete multi-modal features where $M$ is the number of modalities. It is known that, for the multi-modal representation, the goal is to obtain the complementary feature and address the modal gap by exploiting multi-modal cues [5, 9, 29]. Thus, we propose a Graph-based Feature Representation (GFR) to fully leverage the complementary information and mitigate modality gaps. Specifically, we construct a multi-relation graph that captures the diverse relationships between the multi-modal features of the different subjects. That is, each node is the multi-modal feature of each subject, and the edge encodes the similarity between subjects of each modality. There are $M$ modalities corresponding to $M$ types of edges. We utilize the cosine function to calculate the similarity $A^{(r)}$ between subjects where $r$ is the $r$-th relation via utilizing the $m$-th modal features. To learn a more effective graph, we introduce an additional regularization constraint $\mathcal{L}_S(A^{(r)})$ into the graph [11, 32]. Additionally, we construct a supervision graph $\hat{A}$ with the corresponding labels of subjects. Thus, the total loss function $\mathcal{L}_G$ is defined as,

$$\mathcal{L}_G(\mathcal{A}) = \frac{1}{M}\sum_{r=1}^{M}\big(\mathcal{L}_S(A^{(r)}) + ||\hat{A} - A^{(r)}||_2^2\big), \tag{1}$$

where $\mathcal{A} = \{A^{(1)} \cdots A^{(M)}\}$ denotes a multi-relation graph. The $\hat{A}$ is utilized to better identify the category of nodes. If the $i$ and $j$ subjects are of the same category, $\hat{A}_{ij} = 1$; otherwise, $\hat{A}_{ij} = 0$. $\|\cdot\|_2$ denotes the Frobenius norm. Based on the above, we then employ the multi-layer Dynamical Graph Convolutional Network (DGCN) [22] to learn the context-aware representation for each subject by capturing the dependencies of different subjects. To be specific, the layer-wise message propagation rule of DGCN is defined as,

$$F^{(r,l+1)} = \mathrm{ReLU}(A^{(r)}F^{(r,l)}\Theta^{(r,l)}), \tag{2}$$

where $l = 0, 1 \cdots L-1$ and $F^{(r,0)} = \mathrm{Concat}[X^{(1)} \cdots X^{(M)}]$ and $\Theta^{(m,l)}$ denotes the layer-wise trainable weight parameters. We fuse the $l$-th layer output as $\hat{F}^{l+1} = \sum_{r=1}^{M} w_r F^{(r,l)}$ to obtain rich multi-modal features where $w_r$ is the learnable parameter. We derive the set of representations $\{\hat{F}^{(1)} \cdots \hat{F}^{(M)}\}$ from the final fused multi-modal features $\hat{F}^L$. Comparing to the original features $X^{(m)}$, the learned representations $\hat{F}^{(m)}$ involves more contextual information.

### 2.2 Prompt-guided Learning Network

To further learn modality-specific representation, we propose a Prompt-guided Learning Network (PLN) to mine the discriminative regions of brain diseases.

**Semantic Prompts:** A set of semantic prompts is generated to provide a better understanding of brain disease via LLM. To be specific, GPT-4 is utilized to yield semantic information of related brain disease based on specific instructions, e.g., "significant hippocampal atrophy: AD subjects often show significant

atrophy of the hippocampal region". This approach provides a level of semantic guidance for lesion regions. Technically, the relevant concepts are expressed as,

$$I_s = \{I_1, I_2 \cdots I_C\} = \text{GPT-4}\{\text{Class}_1, \text{Class}_2 \cdots \text{Class}_C\}, \tag{3}$$

where Class refers to the name of the disease category and $C$ denotes the number of category. Next, the frozen text encoder CLIP is employed on concept prompts $I_s$ to obtain semantic features $X_s \in \mathbb{R}^{C\times D}$ of all categories.

**Cross-Attention Module:** We introduce the Cross-Attention Module (CAM) to boost interactive learning among different modalities [32]. Specifically, we first transform $\hat{F}^{(m)}$ into input query $Q^{(m)} = (q_1^{(m)} \cdots q_N^{(m)})$, key $K^{(m)} = (k_1^{(m)} \cdots k_N^{(m)})$ and value $V^{(m)} = (v_1^{(m)} \cdots v_N^{(m)})$ respectively by using three different linear projections where $m = 1 \cdots M$ and $N$ is the number of subjects. Then, the features of each modality can be enhanced by aggregating the message from other modalities. This process can be achieved as follows,

$$T_i^{(m,n)} = \frac{\exp[(q_i^{(m)})^T k_i^{(n)}/\tau]}{\sum_{n=1}^{M} \exp[(q_i^{(m)})^T k_i^{(n)}/\tau]}, \hat{v}_i^{(m)} = \sum_{n=1}^{M} T_i^{(m,n)} v_i^{(n)} + \alpha v_i^{(n)}, \tag{4}$$

where $T_i$ represents cross-modal affinity matrix for the $i$-th subject, and $\tau$ is the scaling factor. $\alpha > 0$ denotes the weight parameter to balance the two terms.

**Prompt-guided Embedding Network (PEN):** It aims to guide semantic prompts to obtain the discriminant local cues of each modality and then embed them into global features to highlight the actual lesion regions. Specifically, we perform the different linear projections on the $\hat{v}_i^{(m)}$ to obtain the global features $\hat{v}_{i,g}^{(m)}$ and local features $\hat{v}_{i,p}^{(m)}$ for the $m$-th modality. Let $\hat{V}_g^{(m)} = \{\hat{v}_{1,g}^{(m)} \cdots \hat{v}_{N,g}^{(m)}\} \in \mathbb{R}^{N\times d}$ and $\hat{V}_p^{(m)} = \{\hat{v}_{1,p}^{(m)} \cdots \hat{v}_{N,p}^{(m)}\} \in \mathbb{R}^{N\times P\times d}$ represent global features and local features of all subjects where $P$ and $d$ are the number of part tokens and the dimension of these tokens, respectively. To explore fine-grained local information, we design the Semantic Prompt Guidance Scheme (SPGS), which first applies semantic cues to the local information to compute similarity as follows,

$$S^{(m)} = \text{Sim}\{W_{pro}(X_s), \hat{V}_p^{(m)}\}, S^{(m)} \in \mathbb{R}^{N\times P\times C}, \tag{5}$$

where $W_{pro}$ denotes the projection matrix. And then this scheme selects the semantic similarity with the highest value as $\hat{S}^{(m)} = \max\{S^{(m)}\} \in \mathbb{R}^{N\times P}$, helping to adjust the weight of the corresponding local information and identifying meaningful lesion regions. Note that we do not know the category of subjects, so we can infer the important lesion areas according to the semantic similarity of the local information to the semantic cue. Thus, the $\hat{S}^{(m)}$ is dotted into local information $\hat{V}_p^{(m)}$ to highlight discriminative regions as $\tilde{V}_p^{(m)}$, enabling the model to focus on relevant lesion regions while suppressing noise from irrelevant areas. Finally, to propagate local information and refine discriminative features, we propose a Relation-aware Embedding Strategy (RES). This strategy integrates fine-grained local information into global contexts by modeling their relationships, deriving the unified representation as follows,

$$Z^{(m)} = \text{Softmax}(\tilde{V}_p^{(m)} \hat{V}_g^{(m)})^T \tilde{V}_p^{(m)} + \hat{V}_g^{(m)}, Z = \text{Concat}[Z^{(1)} \cdots Z^{(M)}]. \tag{6}$$

We then feed $Z$ to a classifier to obtain the predicted labels $\hat{Y}$. We train the whole network in an end-to-end way. The overall loss involves both multi-relation graph learning loss and label prediction loss, which is formulated as follows,

$$\mathcal{L} = \lambda \mathcal{L}_G(\mathcal{A}) + \eta \mathcal{L}_{Label}(Y, \hat{Y}), \tag{7}$$

where $\mathcal{L}_G$ is defined in Eq.(1) and $\mathcal{L}_{Label}$ denotes the cross-entropy loss function. $Y$ is the corresponding ground-truth labels. $\lambda$ and $\eta$ are set to 1 in experiments.

## 3 EXPERIMENTAL RESULTS AND ANALYSIS

### 3.1 Dataset and Implementation Detail

**Dataset. TADPOLE:** TADPOLE is a subset of the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [19, 20], which contains multi-modal data from patients. Following the setting in [32], we select 598 subjects with six modalities. **ABIDE:** Autism Brain Imaging Data Exchange (ABIDE) [21] is a public dataset for autism research. Following the classical setting in [4], we preprocess this dataset to obtain 871 subjects with four modalities [4, 7]. Table 1 summarizes the subject statistics for both datasets. MMSE and MoCA are two widely used cognitive function screening scales in clinical diagnosis.

**Implementation Detail.** Our model is trained on an NVIDIA GeForce RTX 3090 GPU using PyTorch 1.12.0. On TADPOLE, the dropout rate is 0, the learning rate is 0.012, the number of hidden layers is 10, and the encoding layer is 2. On ABIDE, the dropout rate is 0.75, the learning rate is 0.0023, the number of hidden layers is 10, and the encoding layer is 4. For both datasets, the number of attention heads is 4, and the weight of original multi-modal features $\alpha = 1$. We use 10-fold cross-validation for robust evaluation, with Accuracy (ACC), Area Under the Curve (AUC), Sensitivity (SEN), and Specificity (SPE) metrics.

**Table 1.** Statistics of subjects on two datasets. F/M denotes Female/Male.

| TADPOLE | Gender(F/M) | Age | MMSE | MoCA |
|---|---|---|---|---|
| NC | 114/95 | 72.81±5.96 | 29.13±1.11 | 25.93±2.45 |
| MCI | 144/171 | 70.87±7.19 | 28.14±1.70 | 23.53±3.10 |
| AD | 30/44 | 73.29±7.97 | 22.82±2.93 | 16.86±5.06 |

| ABIDE | Gender(F/M) | Age | Open/Closed-Eye |
|---|---|---|---|
| NC | 90/378 | 16.84±7.23 | 321/147 |
| ASD | 54/349 | 17.07±7.95 | 288/115 |

### 3.2 Comparison Results

We compare our model with 11 state-of-the-art methods of disease prediction. Specifically, we apply MLP [26], which shows potential relative to other complex models in predicting disease. Besides, InceptionGCN [13], PopGCN [23], EV-GCN [8], LGL [3], MMGL [32], and MAFGN [30] are single-graph-based methods; LSTMGCN [12], Multi-GCN [14], MMKGL [18] and MGDR [9] are multiple-graphs-based methods. Compared to these methods, we propose the

**Table 2.** Comparisons 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 [26] | 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 [23] | 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 [12] | 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 [8] | 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 [3] | 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 [32] | 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 [30] | 92.80±0.92 | 93.32±2.10 | – | – | – | – |
| MMKGL [18] | – | – | 91.08±0.59 | 91.01±0.63 | **91.97±0.64** | 90.05±1.37 |
| MGDR [9] | 93.64±3.90 | 94.89±2.96 | 91.39±2.00 | 91.25±2.07 | 89.33±4.55 | 93.16±3.27 |
| **SPromptGL** | **94.81±2.95** | **95.59±2.57** | **92.30±2.58** | **92.22±2.57** | 91.08±3.53 | **93.36±3.67** |

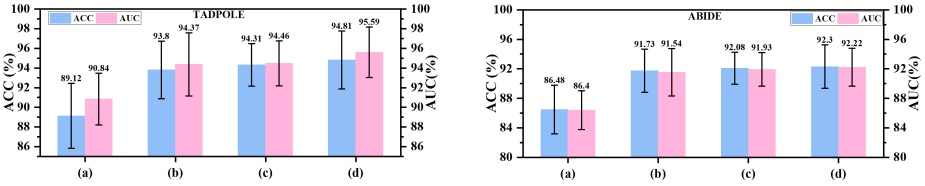


**Fig. 2.** Ablation study results on both datasets. (a) Baseline, (b) Baseline + GFR, (c) Baseline + GFR + RES, (d) Baseline + GFR + PEN (RES + SPGS).

SPromptGL approach, which captures discriminative regions across modalities while enhancing their interaction and fusion. These results are presented in TABLE 2. The proposed method achieves superior performance, with ACC and AUC improvements of 1.17% and 0.7% on TADPOLE, and consistently outperforms the second-best on ABIDE. These results demonstrate the importance of modal-specific discriminative regions for brain disease prediction.

### 3.3 Model Analysis

**Ablation Study:** We construct variants of our method to validate the effectiveness of the proposed modules on two datasets. To be specific, our baseline model only utilizes the multi-modal features from the Cross-Attention Module (CAM) as shown in Fig. 2 (a). (b) We add GFR to learn the rich feature representation instead of directly obtaining the feature representation through CMA. Experimental results show that GFR improves all metrics by approximately 4% on both datasets, demonstrating its ability to effectively capture cross-modal relationships. (c) We integrate RES into our method to embed fine-grained local information into a global context, enabling local message propagation. This outperforms using only GFR, indicating that the features learned are more effective for disease prediction. (d) We introduce SPGS into RES to construct PEN, which utilizes semantic cues to enhance meaningful lesion regions and reduce the interference of non-lesion areas. This leads to consistent improvements on both

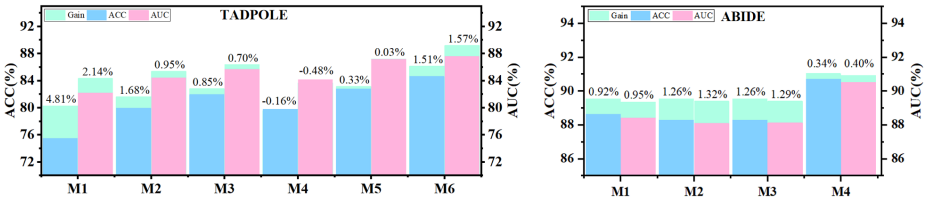


**Fig. 3.** ACC and AUC are obtained for each modality with/without PEN. The green color shows the increased value of each modality with PEN and M* is the *-th modality.

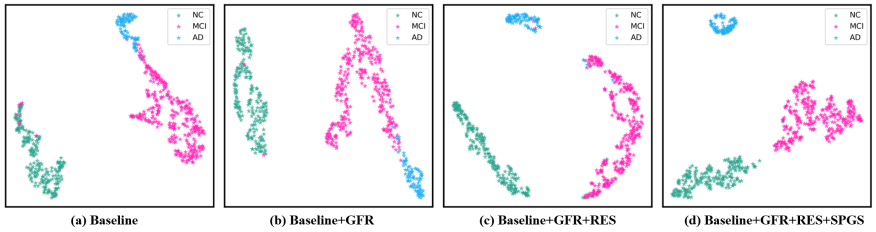


**Fig. 4.** t-SNE visualization of the modality-specific features on TADPOLE.

datasets. Therefore, these results validate that the proposed components play positive roles in enabling better excavation features from different modalities.

To further verify the effectiveness of the proposed PEN, we conduct an in-depth modal ablation analysis. These results show that the performance of each modality is significantly improved after using PEN, as shown in Fig. 3. This fully proves that our method can effectively extract discriminative modal-specific feature representations, enhancing the overall performance of multi-modal learning. It is found that the performance of M4 on TADPOLE decreases slightly, which may be attributed to its insufficient sensitivity to the discriminative regions.

**2D t-SNE Visualization:** To evaluate the effectiveness of our model, we visualize the learned features by incrementally adding the proposed components, as shown in Fig. 4 (a)-(d) on TADPOLE. This visualization has demonstrated success in classifying subjects. Specifically, subjects diagnosed with AD are accurately distinguished from those classified as NC or MCI, and MCI cases are correctly classified without misclassification as AD or NC, ensuring high diagnostic accuracy in Fig. 4 (d). Moreover, it can be found that our results exhibit larger inter-class distances and smaller intra-class distances, indicating that our model is better at capturing differences between subjects of different classes.

## 4 Conclusion

In this paper, we propose to develop a novel prompt-guided graph learning network for a multi-modal brain disease prediction problem. We first construct a relation graph for each modality and then optimize it through a novel graph

learning loss function. Then, a novel multi-layer graph convolutional neural network is used to learn context-enriched feature representation for each subject. After that, the generated semantic prompts are guided to the fine-grained local information to seek out discriminative lesions. Finally, to achieve local information propagation and refine discriminative features, we embed the above fine-grained local information into global contexts by considering their relationship to highlight relevant regions. Experiments on two standard benchmark datasets demonstrate that the proposed approach can achieve superior performance.

**Acknowledgments.** This work was supported by the National Natural Science Foundation of China (62372003), the Natural Science Foundation of Anhui Province (2308085Y40, 2408085J037), and the Key Technologies R&D Program of Anhui Province (202423k09020039).

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

## References

1. Bajestani, G.S., Behrooz, M., Khani, A.G., Nouri-Baygi, M., Mollaei, A.: Diagnosis of autism spectrum disorder based on complex network features. Computer Methods and Programs in Biomedicine **177**, 277–283 (2019)
2. Bzdok, D., Thieme, A., Levkovskyy, O., Wren, P., Ray, T., Reddy, S.: Data science opportunities of large language models for neuroscience and biomedicine. Neuron **112**(5), 698–717 (2024)
3. Cosmo, L., Kazi, A., Ahmadi, S.A., Navab, N., Bronstein, M.: Latent-graph learning for disease prediction. In: Medical Image Computing and Computer Assisted Intervention. vol. 12262, pp. 643–653. Springer (2020). https://doi.org/10.1007/978-3-030-59713-9_62
4. Craddock, C., *et al.*: The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. Frontiers in Neuroinformatics **7** (2013)
5. Deng, Y., Chen, Z., Li, C., Tang, J.: Uncertainty-aware coarse-to-fine alignment for text-image person retrieval. Visual Intelligence **3**, Article no. 6 (2025)
6. Feng, Y., Xu, X., Zhuang, Y., Zhang, M.: Large language models improve alzheimer's disease diagnosis using multi-modality data. pp. 61–66 (2023). https://doi.org/10.1109/MedAI59581.2023.00016
7. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J.: Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping **2**(4), 189–210 (1994)
8. Huang, Y., Chung, A.C.S.: Edge-variational graph convolutional networks for uncertainty-aware disease prediction. In: Medical Image Computing and Computer Assisted Intervention. pp. 562–572. Springer (2020). https://doi.org/10.1007/978-3-030-59728-3_55
9. Jiang, B., *et al.*: Mgdr: Multi-modal graph disentangled representation for brain disease prediction. In: Medical Image Computing and Computer Assisted Intervention. pp. 302–312. Springer (2024). https://doi.org/10.1007/978-3-031-72069-7_29

10. Jiang, H., Cao, P., Xu, M., Yang, J., Zaïane, O.: Hi-gcn: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. Computers in Biology and Medicine **127**, 104096 (2020)
11. Kalofolias, V.: How to learn a graph from smooth signals. preprint arXiv:1601.02513 (2016)
12. Kazi, A., *et al.*: Graph convolution based attention model for personalized disease prediction. In: Medical Image Computing and Computer Assisted Intervention. pp. 122–130. Springer (2019). https://doi.org/10.1007/978-3-030-32251-9_14
13. Kazi, A., *et al.*: Inceptiongcn: Receptive field aware graph convolutional network for disease prediction. In: Information Processing in Medical Imaging. vol. 11492, pp. 73–85. Springer (2019). https://doi.org/10.1007/978-3-030-20351-1_6
14. Kazi, A., krishna, S., 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 (2019). https://doi.org/10.1109/ISBI.2019.8759274
15. Kipf, T., Welling, M.: Semi-supervised classification with graph convolutional networks. preprint arXiv:1609.02907 (2016)
16. Ktena, S.I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., Rueckert, D.: Metric learning with spectral graph convolutions on brain connectivity networks. NeuroImage **169**, 431–442 (2018)
17. LeCun, Y., Bengio, Y.: Convolutional networks for images, speech, and time series, pp. 255–258. MIT Press (1998)
18. Mao, J., Liu, J., Lin, H., Kuang, H., Pan, Y.: Multi-modal multi-kernel graph learning for autism prediction and biomarker discovery. preprint arXiv.2303.03388 (2023)
19. Marinescu, R., *et al.*: The alzheimer's disease prediction of longitudinal evolution (tadpole) challenge: Results after 1 year follow-up. Machine Learning for Biomedical Imaging **019**, 1–60 (2021)
20. Marinescu, R., Oxtoby, N., Young, A., Bron, E., Toga, A., Weiner, M., Barkhof, F., Fox, N., Klein, S., Alexander, D.: Tadpole challenge: Prediction of longitudinal evolution in alzheimer's disease. preprint arXiv.1805.03909 (2018)
21. 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**(2), 659–667 (2014)
22. Miao, X., *et al.*: Lasagne: A multi-layer graph convolutional network framework via node-aware deep architecture. IEEE Transactions on Knowledge and Data Engineering **35**(2), 1721–1733 (2023)
23. Parisot, S., *et al.*: Spectral graph convolutions for population-based disease prediction. In: Medical Image Computing and Computer Assisted Intervention. pp. 177–185. Springer (2017)
24. Peng, L., Cai, S., Wu, Z., Shang, H., Zhu, X., Li, X.: Mmgpl: Multimodal medical data analysis with graph prompt learning. Medical Image Analysis **97**, 103225 (2024)
25. Schuster, M., Paliwal, K.: Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing **45**(11), 2673–2681 (1997)
26. Tolstikhin, I.O., *et al.*: Mlp-mixer: An all-mlp architecture for vision. vol. 34, pp. 24261–24272. Curran Associates, Inc. (2021)
27. Wang, X., *et al.*: Pre-training on high-resolution x-ray images: An experimental study. Visual Intelligence **3**, Article no. 8 (2025)

28. Wen, J., *et al.*: Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation. Medical Image Analysis **63**, 101694 (2020)
29. Xie, X., Cui, Y., Tan, T., Zheng, X., Yu, Z.: Fusionmamba: dynamic feature enhancement for multimodal image fusion with mamba. Visual Intelligence **2**, Article no. 37 (2024)
30. 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 (2023)
31. Zhang, X., Jie, B., Wang, J.: Convolutional recurrent neural network with multiscale kernels on dynamic connectivity network for ad classification. pp. 69–74. Association for Computing Machinery (2023)
32. Zheng, S., *et al.*: Multi-modal graph learning for disease prediction. IEEE Transactions on Medical Imaging **41**(9), 2207–2216 (2022)
33. Zhou, H., Zhang, D.: Graph-in-graph convolutional networks for brain disease diagnosis. In: International Conference on Image Processing. pp. 111–115 (2021). https://doi.org/10.1109/ICIP42928.2021.9506259