MICCAI

Self-guided Knowledge-injected Graph Neural Network for Alzheimer's Diseases

Zhepeng Wang[1], Runxue Bao[2], Yawen Wu[3], Guodong Liu[4], Lei Yang[1], Liang Zhan[3], Feng Zheng[5], Weiwen Jiang[1], and Yanfu Zhang[6(✉)]

[1] George Mason University, Fairfax VA 22032, USA
[2] GE Healthcare, Bellevue WA 98004, USA
[3] University of Pittsburgh, Pittsburgh PA 15260, USA
[4] University of Maryland, College Park MD 20742, USA
[5] Southern University of Science and Technology, Shenzhen GD 518055, CHN
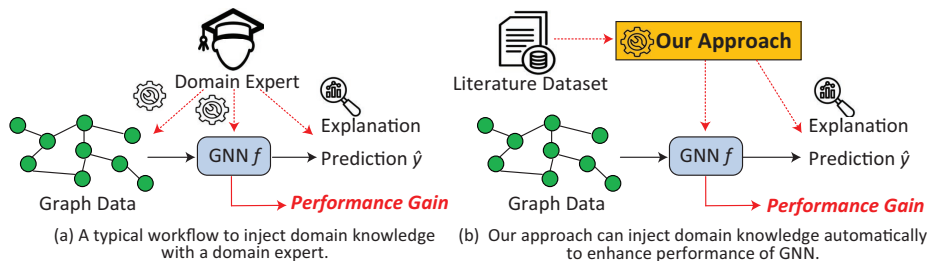[6] William and Mary, Williamsburg VA 23185, USA
`yzhang105@wm.edu`

Abstract. Graph neural networks (GNNs) are proficient machine learning models in handling irregularly structured data. Nevertheless, their generic formulation falls short when applied to the analysis of brain connectomes in Alzheimer's Disease (AD), necessitating the incorporation of domain-specific knowledge to achieve optimal model performance. The integration of AD-related expertise into GNNs presents a significant challenge. Current methodologies reliant on manual design often demand substantial expertise from external domain specialists to guide the development of novel models, thereby consuming considerable time and resources. To mitigate the need for manual curation, this paper introduces a novel self-guided knowledge-infused multimodal GNN to autonomously integrate domain knowledge into the model development process. We propose to conceptualize existing domain knowledge as natural language, and devise a specialized multimodal GNN framework tailored to leverage this uncurated knowledge to direct the learning of the GNN submodule, thereby enhancing its efficacy and improving prediction interpretability. To assess the effectiveness of our framework, we compile a comprehensive literature dataset comprising recent peer-reviewed publications on AD. By integrating this literature dataset with several real-world AD datasets, our experimental results illustrate the effectiveness of the proposed method in extracting curated knowledge and offering explanations on graphs for domain-specific applications. Furthermore, our approach successfully utilizes the extracted information to enhance the performance of the GNN.

Keywords: Alzheimer's Disease · Graph Neural Network · Multimodal.

1 Introduction

Graph neural networks (GNNs) have emerged as robust and versatile algorithms for diverse machine learning tasks involving graph or irregularly structured data,

(a) A typical workflow to inject domain knowledge with a domain expert.

(b) Our approach can inject domain knowledge automatically to enhance performance of GNN.

Fig. 1. Illustration of designing AD-specific GNNs.

as evidenced by a plethora of studies [10, 21, 27, 6, 7, 2]. In recent years, there has been a discernible surge in research endeavors aimed at extending the applicability of GNNs towards addressing broader scientific and health-related challenges, such as drug prediction [26, 14], precision medicine [24, 22], and protein structure prediction [25, 19].

Despite the inherent capability of GNNs to capture crucial structural information within graphs, the direct application of generic GNNs to Alzheimer's Disease (AD) research is not straightforward. This is due to the distinctive properties of brain connectomes, such as fixed node numbers and orders of Regions of Interest (ROIs), which are not typically incorporated in standard GNN architectures. The pursuit of this research direction mandates a collaborative effort between neuroscientists and computer scientists to tailor GNNs specifically for AD, as shown in Fig. 1(a). However, the imperative for frequent interaction with domain experts contrasts with the overarching goal of minimizing manual intervention in machine learning algorithm development. The efficacy of the model may be constrained by the expertise and knowledge of the consulted experts, while the training of reliable domain experts incurs significant time and cost. Moreover, the scarcity of available experts poses further challenges, hindering model development efforts. Furthermore, the opaque nature of GNNs complicates the elucidation of decision-making processes, limiting the assistance that human experts can provide. Besides, although there exist works incorporating human knowledge with graph data in recommendation systems [13, 16] in recent years, they cannot be directly extended to the domain of AD.

In an effort to reduce reliance on expert guidance in the development of GNNs for AD, our objective is to leverage self-collected, coarse, and uncurated domain knowledge instead of relying solely on meticulously curated expert knowledge. This approach raises several pertinent research questions: (1) How can uncurated domain knowledge, readily accessible through sources such as peer-reviewed publications on the internet, be effectively utilized by GNNs? (2) How can this knowledge be identified, curated, and tailored to suit specific datasets and GNN architectures? (3) How can such knowledge be effectively integrated into the model design to enhance performance and interoperability?

To address these challenges, we propose a self-guided multimodal approach to enhance graph representation learning for AD, as depicted in Fig. 1(b). Inspired

by retrieval-augmented generation (RAG) [12, 1, 8, 15] techniques for foundation models, we augment GNNs with extensive domain knowledge extracted from a self-built literature dataset, employing an end-to-end methodology to automatically discern graph-wise and knowledge-wise importance for guiding interpretable predictions. Initially, we integrate multimodal inputs (i.e., graph and natural language) and pretrain the multimodal GNN. Subsequently, we compute the importance scores of brain substructures and external knowledge, represented by explainable mask values. Finally, we refine the pretrained model through augmented brain connectomes guided by the masks to enhance performance. Our method contributions can be summarized as follows:

* We propose an approach to automatically incorporate domain knowledge into GNNs to minimize human intervention, leveraging an external source of coarse and uncurated knowledge, and introducing a novel "soft" augmented retrieval generation method for integration with relevant subjects.
* Our approach aims to concurrently enhance the utility and interpretability of GNNs by identifying graph-wise and knowledge-wise explanations relevant to predictions. Moreover, the extracted explanations are tailored to the given dataset and GNNs thereby guiding graph augmentation for model fine-tuning.
* Experimental results on representative AD datasets demonstrate that our method can effectively inject domain knowledge into GNNs, thereby improving performance and facilitating the generation of customized explanations.

2 Method

Overview. To integrate external knowledge into GNNs, we will fuse the embeddings from graphs and external knowledge guided by their relevance to the prediction. More specifically, we learn a pair of masks to characterize the relevance and use the masked embeddings to fine-tune the prediction model.

We first formulate a multimodal GNN denoted as f, which accepts brain connectome data g alongside natural language data $\mathcal{K}$ as inputs. In essence, this entails augmenting canonical GNNs with external knowledge data. Subsequently, we undertake the pretraining of f using gathered data. We then identify significant substructures within brain connectomes and pertinent knowledge relevant to prediction tasks. These are delineated by real-valued masks parameterized by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. By computing the masks jointly, we facilitate a "soft" retrieval mechanism from the knowledge data, wherein the masks function as explanations for both graph-wise and knowledge-wise aspects of predictions made by f. Ultimately, we refine the parameters of f through a process of fine-tuning, leveraging graph augmentation in conjunction with the "soft" retrieval outcomes. Specifically, we utilize the computed masks to guide edge sampling, thereby preserving critical information while augmenting data diversity. The design overview is shown in Fig. 2 and further elucidation of our approach follows.

Multimodal GNN equipped with external knowledge. We propose the integration of canonical GNNs with external knowledge, culminating in a multimodal GNN. Specifically, our multimodal GNN comprises a backbone GNN
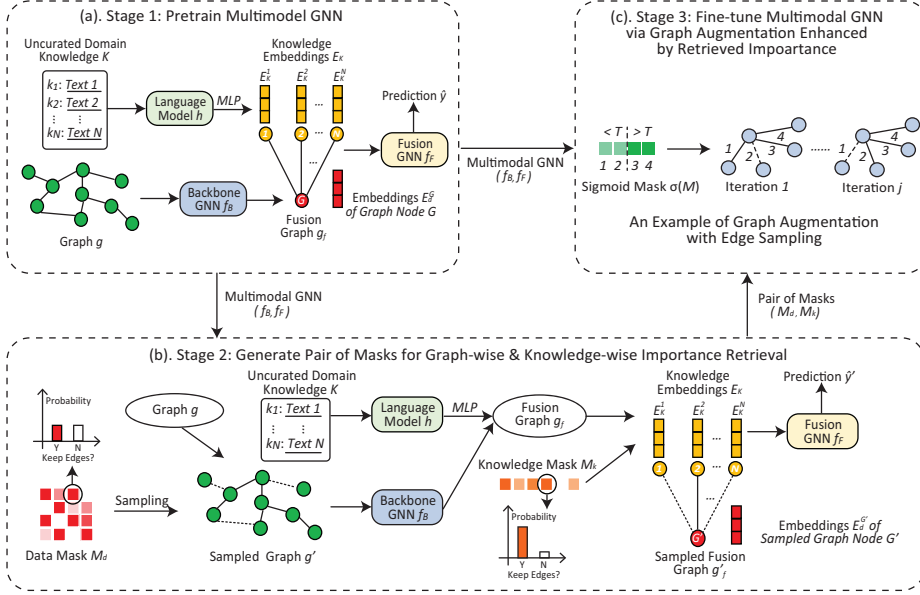
Fig. 2. Design overview of our approach.

denoted as f_B and a fusion GNN denoted as f_F. Given the graph data g, the backbone GNN f_B is employed to generate a node embedding $E_d^G = f_B(g)$. The uncurated domain knowledge $\mathcal{K}$ is represented as a collection of language sequences, denoted as $\mathcal{K} = \{k_1, k_2, ..., k_N\}$, where each knowledge item k_i corresponds to a summarization of domain knowledge. For instance, the abstract of a paper focusing on Alzheimer's Disease (AD) can be considered as a language sequence k_i within the uncurated domain knowledge $\mathcal{K}$ for AD. To integrate $\mathcal{K}$ with E_d^G, we compute $E_\mathcal{K}^i = MLP(h(k_i)), \forall k_i \in \mathcal{K}$, where a pretrained language model h followed by a multi-layer perceptron MLP is utilized to generate the language embedding. We denote the knowledge embeddings collectively as $E_\mathcal{K} = \{E_\mathcal{K}^i | i = idx(k_i), k_i \in \mathcal{K}\}$ with $|\mathcal{K}| = N$. Utilizing E_d^G and $E_\mathcal{K}$, a fusion graph g_f is constructed by connecting node G with feature E_d^G to each node i in $\mathcal{K}$ with feature $E_\mathcal{K}^i$. This allows us to modify cross-modal edges to represent retrieval. The final prediction $\hat{y}$ for the target task is generated by the fusion GNN f_F, where $\hat{y} = f_F(g_f)$. Subsequently, following the forward propagation of the multimodal GNN f as defined, the model can be trained end-to-end using multimodal inputs g and $\mathcal{K}$, leveraging any arbitrary loss function $\mathcal{L}$ specific to the target task. The design of our multimodal GNN is illustrated in Fig. 2(a).

Graph-wise and knowledge-wise importance retrieval using masks. Given the pretrained GNN, we aim to identify the importance score represented by real-value masks of substructures in the graphs and the retrieved knowledge, to indicate the crucial components contributing to the prediction, and the details are shown in Fig. 2(b). Following the fashion in previous works, we define the

explanation of predictions on the multimodal input g and $\mathcal{K}$ as a subgraph $\tilde{g}$ and a subset $\tilde{\mathcal{K}}$. Since directly sampling them from g and $\mathcal{K}$ is computationally prohibitive and indifferentiable, we parameterized the sampling process as two types of learnable masks combined with Gumbel Softmax function [9], i.e., data mask M_d parameterized with $\boldsymbol{\alpha}$ on the graph g and knowledge mask M_k parameterized with $\boldsymbol{\beta}$ on the fusion graph g_f. More specifically, for a given mask M (M_d or M_k) parameterized with $\boldsymbol{\phi}$ ($\boldsymbol{\alpha}$ for M_d or $\boldsymbol{\beta}$ for M_k), Gumbel Softmax function is applied to arbitrary entry within M as $m_{i,j} = \frac{\exp\left[(\phi_{i,j}+g_{i,j})/\tau\right]}{\exp\left[(\phi_{i,j}+g_{i,j})\right]+\exp\left(g'_{i,j}\right)}$, where $\phi_{i,j}$ denotes the entry value indexed with (i,j) in M. $g_{i,j}$ and $g'_{i,j}$ are two independent random noise sampled from Gumbel distribution $Gumbel(0,1)$ controled by temperature τ.

Given graph $g = (V_g, \boldsymbol{W}_g)$, where V_g denotes the node set of g and $\boldsymbol{W}_g \in \mathbb{R}^{|V_g|\times|V_g|}$ represents the weighted adjacency matrix of g, the matrix M'_d sampled from M_d is applied to $\boldsymbol{W}_g$ to approximate the sampling of edges as $\boldsymbol{W}'_g = M'_d \odot \boldsymbol{W}_g$, where $\odot$ is the element-wise matrix multiplication. With $\boldsymbol{W}'_g$, we sample $g' = (V_g, \boldsymbol{W}'_g)$, and calculate the node embedding as $E_d^{G'} = f_B(g')$. Node embedding $E_d^{G'}$ is then used together with the knowledge embedding $E_{\mathcal{K}}$ to construct the fusion graph $g_f = (V_f, \boldsymbol{A}_f)$, where V_f denotes the node set of g_f and $\boldsymbol{A}_f \in \mathbb{R}^{|V_f|-1}$ represents the unweighted adjacency matrix of g_f, where all the entries are one. Similarly, for fusion graph g_f, we apply the sampled mask M'_k to $\boldsymbol{A}_f$ to approximate the sampling of edges as $\boldsymbol{A}'_f = M'_k \odot \boldsymbol{A}_f$ and get the sampled fusion graph g'_f. The final prediction is calculated by $f_F(g'_f)$.

We define $\mathcal{L}_{exp}$ to learn the masks with hyperparameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4$,

$$\mathcal{L}_{exp} = \lambda_1 \mathcal{L}_{mask} + \lambda_2 \mathcal{L}_{clf} + \lambda_3 \mathcal{L}_{spas} + \lambda_4 \mathcal{L}_{disc}. \tag{1}$$

$\mathcal{L}_{mask} = -\sum_{c=1}^{C} \mathbb{1}[\hat{y} = c] \log P_f(\hat{y}' = \hat{y}|g', g'_f)$ denotes the disagreement between the prediction $\hat{y}$ made on the original inputs and the prediction $\hat{y}'$ made on the sampled inputs. C is the class number in the target task. $\mathcal{L}_{clf} = -\sum_{c=1}^{C} \mathbb{1}[y = c] \log P_f(\hat{y}' = y|g', g'_f)$ encourages the consistency between $\hat{y}'$ and the label y. $\mathcal{L}_{spas} = \frac{1}{|M_d|}\sum_{i,j}\sigma(M_d^{i,j}) + \frac{1}{|M_k|}\sum_{i,j}\sigma(M_k^{i,j})$ controls the sparsity of the two masks, where $\sigma(\cdot)$ represents the sigmoid function. $|M_d|$ and $|M_k|$ are the number of entries in M_d and M_k, respectively. $\mathcal{L}_{disc} = \sum_{i,j}\frac{\mathcal{L}_{ent}(\sigma(M_d^{i,j}))}{|M_d|} + \sum_{i,j}\frac{\mathcal{L}_{ent}(\sigma(M_k^{i,j}))}{|M_k|}$ promotes the discreteness of the two masks, where $\mathcal{L}_{ent}(\cdot)$ is the binary entropy function.

This section only considers one pair of M_d and M_k for simplicity. However, we can generalize it to multiple pairs cases. For instance, if genders are provided, we can define $M^{female} = \{M_d^{female}, M_k^{female}\}$ and $M^{male} = \{M_d^{male}, M_k^{male}\}$, and train each pair of masks separately.

Graph augmentation enhanced by retrieved importance. We propose a graph augmentation method guided by the crucial components in prediction, i.e., the learned masks, to fine-tune the pretrained multimodal GNN f, aiming to enhance model performance. Specifically, during fine-tuning, we employ a threshold T on $\sigma(M)$ to guide edge sampling as follows: for any arbitrary entry

value m_i, we retain the corresponding edge when $m_i \geq T$; otherwise, we randomly sample the corresponding edge with a probability of 0.5 for each iteration. A schematic illustration of this process is provided in Fig. 2(c), where m_1 and m_2 are both less than T, while m_3 and m_4 exceed T. Consequently, edges 3 and 4 are retained across both iteration 1 and j, whereas edge 1 exists solely in iteration 1, and edge 2 appears only in iteration j. This methodology can be uniformly applied to all computed masks, thereby substantially increasing input diversity to f concerning less significant edges, while retaining crucial edges for prediction, as indicated by the entry values within the masks.

3   Experiments

Datasets and Settings. We evaluate our approach on two AD datasets, i.e., OASIS [11] and ADNI-D [17]. For each, we include two graph datasets in different modalities, where one is derived from DTI imaging while the other one is derived from fMRI imaging.

* **OASIS** [1] contains 815 subjects. 155 subjects are diagnosed with AD and the others are seronegative controls. For each subject, there are 132 regions of interest (ROIs) based on Harvard-Oxford Atlas [4] and AAL Atlas [20]. There are 459 females, including 66 AD patients and 393 seronegative controls, and 356 males, including 89 AD patients and 267 seronegative controls.
* **ADNI-D** [2] contains 340 subjects. 154 patients are diagnosed with mild cognitive impairment (MCI), which is the early stage of AD. The others are seronegative controls. For each subject, 85 ROIs are derived from T1-weighted MRI using FreeSurfer (V6.5) [3]. There are 210 females, including 84 MCI patients and 126 seronegative controls, and 130 males, including 70 MCI patients and 60 seronegative controls.

For domain knowledge $\mathcal{K}$ of AD, we collected 20,108 records related to AD in the last 20 years from PubMed [3]. Each record consists of the title and the abstract of a paper. We use BERT-Large [5] to encode $\mathcal{K}$. BERT-Large is pretrained with whole word masking on the uncased version of the training corpus.

We evaluate GCN [10], GINE [27], and GAT [21] as the architecture of both the backbone and fusion GNN within the multimodel GNN (MM-GNN). Given the prior knowledge about the gender distribution of the dataset, we introduce two pairs of masks for graph-wise and knowledge-wise importance retrieval, where one is for males and the other is for females. Please refer to the supplementary materials for training details.

Main Results. The proposed MM-GNN significantly outperforms the plain version of GNN, and the MM-GNN with fine-tuning further improves the performance. Table 1 reports the main results for the evaluation of our method. As

---

[1] https://sites.wustl.edu/oasisbrains/home/oasis-3/
[2] https://adni.loni.usc.edu/
[3] https://pubmed.ncbi.nlm.nih.gov/

Table 1. Main results.

Method	OASIS (DTI)			OASIS (fMRI)			ADNI-D (DTI)			ADNI-D (fMRI)		
	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
GCN	0.6466	0.5246	0.4840	0.6245	0.5376	0.5003	0.5765	0.4642	0.4619	0.5618	0.4978	0.5301
MM-GCN (ours)	0.7141	0.5220	0.5639	0.7387	0.5774	0.5962	0.6294	0.5629	0.6133	0.6235	0.5852	0.6111
MM-GCN-F (ours)	**0.7362**	**0.5381**	**0.5805**	**0.7436**	**0.5774**	**0.5989**	**0.6471**	**0.5821**	**0.6406**	**0.6441**	**0.5936**	**0.6367**
GINE	0.7448	0.5238	0.5332	0.6405	0.5476	0.5251	0.5765	0.5966	0.5242	0.5676	0.5862	0.5439
MM-GINE (ours)	0.7521	**0.5777**	0.6004	**0.7558**	0.5641	0.5939	0.6706	0.6286	0.6610	**0.6765**	0.6450	**0.6671**
MM-GINE-F (ours)	**0.7926**	0.5733	**0.6245**	0.7534	**0.6030**	**0.6121**	**0.6853**	**0.6720**	**0.6774**	0.6647	**0.6716**	0.6556
GAT	0.5975	0.5340	0.4925	0.6847	0.4809	0.5184	0.5529	0.5300	0.4617	0.5824	0.4994	0.4512
MM-GAT (ours)	0.7436	**0.6098**	0.5983	0.7632	0.5828	0.6040	0.6588	0.6164	0.6354	0.6559	0.6072	0.6252
MM-GAT-F (ours)	**0.7460**	0.6085	**0.6022**	**0.7779**	**0.6008**	**0.6143**	**0.6706**	**0.6204**	**0.6557**	**0.6618**	**0.6102**	**0.6446**
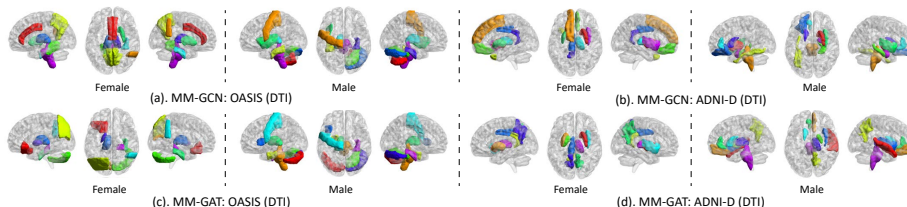
Fig. 3. Brain saliency maps identified by our approach. Top 10 salient ROIs are highlighted.

shown in Table 1, the performance can be improved significantly when the domain knowledge is injected into the inference of GNN. For example, on the two datasets of ADNI-D, compared with vanilla GCN, 10.08%, 19.41% and 24.03% relative improvement is gained in ACC, AUC and F1 on average, respectively. Moreover, the performance can be further enhanced when fine-tuning the multimodal GNN guided by the generated masks. For instance, on the two datasets of ADNI-D, compared with plain MM-GCN, 3.06%, 2.43% and 4.32% relative improvement can be achieved in ACC, AUC and F1 on average, respectively.

Qualitative Analysis on Graph-wise Masks. To analyze the generated graph-wise masks, we calculate the importance score of each ROI based on the graph-wise mask and highlight the top 10 salient ROIs on the brain saliency maps in Fig. 3. Note that the color in each brain saliency map of Fig. 3 is applied to distinguish different ROIs only. As shown in Fig. 3, it is clear that the salient ROIs are distinct for males and females, which is consistent with the existing studies [18, 23]. Moreover, MM-GNNs with different architectures can also show a distinct preference for their prediction even if they are trained on the same dataset. For instance, MM-GCN and MM-GAT share 5 salient ROIs and have 5 unique salient ROIs. Such kinds of insights are usually difficult for human experts to extract but can be automatically captured by our approach.

Qualitative Analysis on Knowledge-wise Masks. Our approach can automatically adapt its way to employ the domain knowledge to the change in the gender, backbone and dataset. Fig. 4 shows the distribution of values within the knowledge-wise masks, which can be interpreted as the distribution of importance scores of domain knowledge. And we have several observations. First,
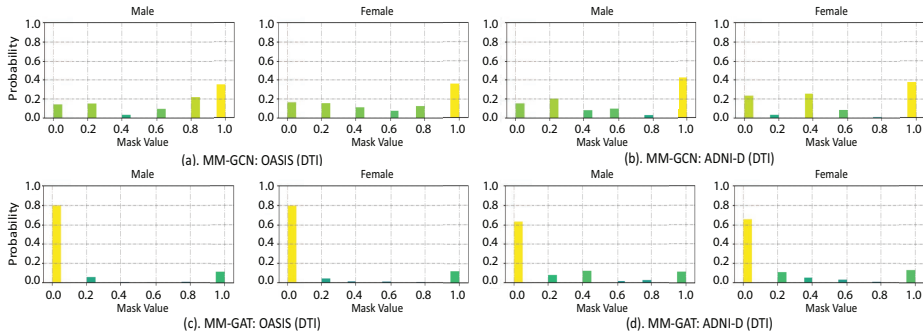
Fig. 4. Distribution of importance scores of domain knowledge from our approach.

Table 2. Ablation study on the size of domain knowledge.

Method	OASIS (DTI)			OASIS (fMRI)			ADNI-D (DTI)			ADNI-D (fMRI)		
	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
MM-GCN (100%)	**0.7141**	0.5220	**0.5639**	**0.7387**	**0.5774**	**0.5962**	**0.6294**	**0.5629**	**0.6133**	**0.6235**	**0.5852**	**0.6111**
MM-GCN (10%)	0.6822	0.5367	0.4852	0.6405	0.5636	0.4921	0.5941	0.5191	0.5833	0.6118	0.5682	0.5656
MM-GCN (1%)	0.6613	**0.5388**	0.4993	0.6429	0.5573	0.5264	0.5912	0.5277	0.5871	0.6235	0.5610	0.5755
MM-GINE (100%)	0.7521	**0.5777**	**0.6004**	0.7558	**0.5641**	**0.5939**	0.6706	**0.6286**	**0.6610**	**0.6765**	0.6450	**0.6671**
MM-GINE (10%)	0.8184	0.5546	0.5115	0.7644	0.5294	0.5801	0.6676	0.6206	0.6479	0.6059	**0.6806**	0.5754
MM-GINE (1%)	**0.8221**	0.5206	0.5152	**0.8221**	0.5078	0.5244	**0.6735**	0.6184	0.6608	0.6206	0.6937	0.5933
MM-GAT (100%)	0.7436	**0.6098**	**0.5983**	0.7632	**0.5828**	**0.6040**	**0.6588**	**0.6164**	**0.6354**	**0.6559**	**0.6072**	**0.6252**
MM-GAT (10%)	0.7485	0.5665	0.5896	0.8184	0.5317	0.5012	0.6088	0.5578	0.5824	0.5559	0.4828	0.5475
MM-GAT (1%)	**0.7718**	0.5562	0.5893	**0.8221**	0.5176	0.5290	0.6324	0.5822	0.5860	0.6118	0.5470	0.5306

males and females have different ways to leverage the given domain knowledge for inference. Second, different architectures of MM-GNN show different patterns on the knowledge-wise mask. Compared with MM-GCN, MM-GAT tends to have a more sparse distribution. Last, even when the domain is the same (i.e., AD), our approach can propose different paradigms to leverage the domain knowledge when the dataset is different.

Ablation Study on the Size of Domain Knowledge. We conducted the ablation study on the size of the domain knowledge, where we randomly sampled 1% and 10% records from the complete set of domain knowledge $\mathcal{K}$ to train the MM-GNN. The results are shown in Table 2 and we have two observations. First, the performance improvement is minor and even suffers from degradation when increasing the size of domain knowledge from about 200 (1%) to about 2,000 (10%). Second, the benefits from a larger set of domain knowledge become obvious when increasing the size of domain knowledge to more than 20,000 (100%). Therefore, we can conclude that although the performance may have little improvement when increasing the set of domain knowledge to a moderate size, a noticeable boost in performance can be observed when increasing the size of domain knowledge to a much larger size.

4 Conclusion

We introduce a self-guided approach to autonomously integrate domain knowledge into GNNs to harness collected uncurated AD knowledge. Our approach can effectively extract curated knowledge and explanations on graphs for AD and guide the fine-tuning to improve the performance of GNNs. Extensive experiments on real-world AD datasets demonstrate the effectiveness of our method.

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

References

1. Borgeaud, S., Mensch, A., Hoffmann, J., Cai, T., Rutherford, E., Millican, K., Van Den Driessche, G.B., Lespiau, J.B., Damoc, B., Clark, A., et al.: Improving language models by retrieving from trillions of tokens. In: International conference on machine learning. pp. 2206–2240. PMLR (2022)
2. Cui, H., Dai, W., Zhu, Y., Li, X., He, L., Yang, C.: 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. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis: I. segmentation and surface reconstruction. Neuroimage **9**(2), 179–194 (1999)
4. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al.: An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. Neuroimage **31**(3), 968–980 (2006)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
6. Gilmer, J., Schoenholz, S.S., et al.: Neural message passing for quantum chemistry. In: International Conference on Machine Learning. pp. 1263–1272. PMLR (2017)
7. Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., Catasta, M., Leskovec, J.: Open graph benchmark: Datasets for machine learning on graphs. Advances in neural information processing systems **33**, 22118–22133 (2020)
8. Izacard, G., Lewis, P., Lomeli, M., Hosseini, L., Petroni, F., Schick, T., Dwivedi-Yu, J., Joulin, A., Riedel, S., Grave, E.: Few-shot learning with retrieval augmented language models. arXiv preprint arXiv:2208.03299 (2022)
9. Jang, E., Gu, S., Poole, B.: Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144 (2016)

10. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
11. LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A.G., et al.: Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. MedRxiv pp. 2019–12 (2019)
12. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.t., Rocktäschel, T., et al.: Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in Neural Information Processing Systems **33**, 9459–9474 (2020)
13. Liang, S., Shao, J., Zhang, J., Cui, B.: Graph-based non-sampling for knowledge graph enhanced recommendation. IEEE Transactions on Knowledge and Data Engineering **35**(9), 9462–9475 (2023)
14. Lin, X., Quan, Z., Wang, Z.J., Ma, T., Zeng, X.: Kgnn: Knowledge graph neural network for drug-drug interaction prediction. In: IJCAI. vol. 380, pp. 2739–2745 (2020)
15. Luo, Z., Xu, C., Zhao, P., Geng, X., Tao, C., Ma, J., Lin, Q., Jiang, D.: Augmented large language models with parametric knowledge guiding. arXiv preprint arXiv:2305.04757 (2023)
16. Lyu, Z., Wu, Y., Lai, J., Yang, M., Li, C., Zhou, W.: Knowledge enhanced graph neural networks for explainable recommendation. IEEE Transactions on Knowledge and Data Engineering **35**(5), 4954–4968 (2022)
17. Mackin, R.S., Insel, P.S., Landau, S., Bickford, D., Morin, R., Rhodes, E., Tosun, D., Rosen, H.J., Butters, M., Aisen, P., et al.: Late-life depression is associated with reduced cortical amyloid burden: Findings from the alzheimer's disease neuroimaging initiative depression project. Biological psychiatry **89**(8), 757–765 (2021)
18. Subramaniapillai, S., Rajagopal, S., Snytte, J., Otto, A.R., Einstein, G., Rajah, M.N., Group, P.A.R., et al.: Sex differences in brain aging among adults with family history of alzheimer's disease and apoe4 genetic risk. NeuroImage: Clinical **30**, 102620 (2021)
19. Tang, H., Ma, G., et al.: Commpool: An interpretable graph pooling framework for hierarchical graph representation learning. Neural Networks **143**, 669–677 (2021)
20. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. Neuroimage **15**(1), 273–289 (2002)
21. Veličković, P., Cucurull, G., et al.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
22. Wang, Y., Wang, Y.G., Hu, C., Li, M., Fan, Y., Otter, N., Sam, I., Gou, H., Hu, Y., Kwok, T., et al.: Cell graph neural networks enable the precise prediction of patient survival in gastric cancer. NPJ precision oncology **6**(1), 45 (2022)
23. Williamson, J., Yabluchanskiy, A., Mukli, P., Wu, D.H., Sonntag, W., Ciro, C., Yang, Y.: Sex differences in brain functional connectivity of hippocampus in mild cognitive impairment. Frontiers in Aging Neuroscience **14**, 959394 (2022)
24. WOLTERINK, J., SUK, J.: Geometric deep learning for precision medicine. KEY ENABLING TECHNOLOGY FOR SCIENTIFIC MACHINE LEARNING **60**
25. Xia, T., Ku, W.S.: Geometric graph representation learning on protein structure prediction. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 1873–1883 (2021)
26. Xiong, J., Xiong, Z., Chen, K., Jiang, H., Zheng, M.: Graph neural networks for automated de novo drug design. Drug Discovery Today **26**(6), 1382–1393 (2021)

27. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? arXiv preprint arXiv:1810.00826 (2018)