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

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`

A Detailed Experimental Settings

Table 1. Specifications of hyper-parameters of the GNN architecture and the language model. The specifications of GNN architecture are shared by all the evaluated GNNs unless explicitly specified. The node features with 16 dimensions are randomly initialized with standard Gaussian distribution. And the random seed is 42, which is used for all the experiments in this work.

GNN Architecture		Language Model Settings	
Hyper-parameters	Value	Hyper-parameters	Value
# GNN Layers	2	Model Architecture	BERT-Large
Hidden Dimension	64	Pretrain Technique	Whole Word Masking
# Attention Head (GAT)	1	Embedding Dimension	1024
Edge Feature Dimension (GINE)	1	# MLP Layers	2
Node Feature Dimension	16	MLP Hidden Dimension	64

Table 2. Specification of hyper-parameters to train the multimodal GNN (MM-GNN) and the pairs of graph-wise and knowledge-wise masks. The specifications are shared by all the evaluated GNNs. For the hyper-parameters with a list of values, it means that we did a grid search over these values and picked the one with the best performance. The best value varies for different experimental settings.

Multimodal GNN (MM-GNN)		Multimodal Masks (Graph & Domain Knowlege)	
Hyper-parameters	Value	Hyper-parameters	Value
Batch Size	16	Batch Size	16
# Epoch	2000	Gumbel Softmax Initial Temperature	5
Optimizer	Adam	Gumbel Softmax Decay Rate	0.998
Weight Decay	5e-4	$\lambda_1, \lambda_2, \lambda_4$	1
Dropout Rate	0.5	λ_3	{0.5, 1, 2.5, 5, 10}
Learning Rate	$\{10^{-2}, 5 \times 10^{-3}, 10^{-3}, 5 \times 10^{-4}, 10^{-4}, 10^{-5}, 10^{-6}\}$	Learning Rate	0.01

Table 3. Specification of hyper-parameters to fine-tune the MM-GNN via graph augmentation with edge sampling using the learned masks. The specifications are shared by all the evaluated GNNs unless explicitly specified. For the hyper-parameters with a list of values, it means that we did a grid search over these values and picked the one with the best performance. The best value varies for different experimental settings.

Graph Augmentation with Edge Sampling			
Hyper-parameters	Value	Hyper-parameters	Value
Batch Size	16	Threshold for Graph	0.8
# Epoch	2000	Threshold for Domain Knowledge	0.5
Optimizer	Adam	Threshold for Graph (GINE)	0.8
Weight Decay	5e-4	Threshold for Domain Knowledge (GINE, DTI)	0.15
Learning Rate	$\{10^{-2}, 5 \times 10^{-3}, 10^{-3}, 5 \times 10^{-4}, 10^{-4}, 10^{-5}, 10^{-6}\}$	Threshold for Domain Knowledge (GINE, fMRI)	0.2

Table 4. Details about how we collected domain knowledge for AD. We first queried PubMed with the listed keywords to obtain the related records about AD in the last 20 years. We then filtered them by ensuring that specific attributes of each record were not empty. We further restrict the length of raw text to less than 420 words to avoid the undesired truncation on the input to the language model. Therefore, we got a set of domain knowledge with 20108 records to be used in all the experiments.

Domain Knowledge for AD	
Starting Date: 04/01/2023	End Date: 03/31/2023
# Length of Record: <420 words	# Records: 20108
Keywords for Query	*Alzheimer's Disease, Subjective Cognitive Impairment, Mild Cognitive Impairment, Dementia, Cognitive Decline*
Non-empty Attributes for Filtering	*First Author, Affiliation, Journal Title, Pub Date, ArticleTitle, AbstractText, KeywordList, MeshHeadingList, Substance* and *ReferenceList*