

BrainPrompt: Multi-Level Brain Prompt Enhancement for Neurological Condition Identification

Jiaxing Xu^{1*} , Kai He^{2*}, Yue Tang³, Wei Li¹, Mengcheng Lan⁴, Xia Dong¹, Yiping Ke¹, and Mengling Feng²

¹ College of Computing and Data Science, Nanyang Technological University, Singapore

jiaxing003@e.ntu.edu.sg

² Saw Swee Hock School of Public Health at National University of Singapore, Singapore

³ Xi'an Jiaotong University, P.R. China

⁴ S-Lab, Nanyang Technological University, Singapore

Abstract. Neurological conditions, such as Alzheimer’s Disease, are challenging to diagnose, particularly in the early stages where symptoms closely resemble healthy controls. Existing brain network analysis methods primarily focus on graph-based models that rely solely on imaging data, which may overlook important non-imaging factors and limit the model’s predictive power and interpretability. In this paper, we present BrainPrompt, an innovative framework that enhances Graph Neural Networks (GNNs) by integrating Large Language Models (LLMs) with knowledge-driven prompts, enabling more effective capture of complex, non-imaging information and external knowledge for neurological disease identification. BrainPrompt integrates three types of knowledge-driven prompts: (1) ROI-level prompts to encode the identity and function of each brain region, (2) subject-level prompts that incorporate demographic information, and (3) disease-level prompts to capture the temporal progression of disease. By leveraging these multi-level prompts, BrainPrompt effectively harnesses knowledge-enhanced multi-modal information from LLMs, enhancing the model’s capability to predict neurological disease stages and meanwhile offers more interpretable results. We evaluate BrainPrompt on two resting-state functional Magnetic Resonance Imaging (fMRI) datasets from neurological disorders, showing its superiority over state-of-the-art methods. Additionally, a biomarker study demonstrates the framework’s ability to extract valuable and interpretable information aligned with domain knowledge in neuroscience. The code is available at <https://github.com/AngusMonroe/BrainPrompt>.

Keywords: Brain Network · Prompt Enhancement · fMRI · Graph Neural Network · Brain Disorder.

* Equal contribution

1 Introduction

Neurological conditions, such as Alzheimer’s disease and Autism, pose significant diagnostic and therapeutic challenges. Early stages are often subtle and difficult to distinguish from healthy controls [16, 17]. Accurate identification and staging are crucial for timely intervention. Brain network analysis using functional magnetic resonance imaging (fMRI) has become a powerful tool to capture functional connectivity and understand brain activity in these conditions [24].

Most existing approaches use graph-based models, especially Graph Neural Networks (GNNs), to encode brain networks derived from neuroimages [14, 31, 27, 26]. However, in clinical practice, diagnosing neurological conditions relies not only on neuroimages but also on various forms of domain knowledge and clinical experience [15]. For example, doctors consider demographic details such as sex, age, and acquisition site to identify sub-population patterns that may influence disease progression. Additionally, clinical observations often take temporal dynamics into account, reflecting how diseases evolve over time [3]. While existing approaches like SEGC [10], SAC-GCN [19], and LG-GNN [31] have attempted to integrate demographic data, they often combine it with imaging data directly as a feature, which overlooks the richer semantic context provided by these factors. This limitation highlights the need for a more nuanced integration of clinical knowledge to improve the accuracy and interpretability of disease classification.

Recent advances in Large Language Models (LLMs) [18, 21] have shown their ability to capture rich, contextual information from text. fTSPL [23] is the first to apply LLMs to brain network analysis by describing the connectivity strength. However, it only considers identity information of regions-of-interest (ROIs). In this paper, we propose BrainPrompt, a novel framework that integrates multi-level knowledge-driven prompts to capture multi-modal information and external knowledge from LLMs for neurological condition identification. BrainPrompt introduces three levels of prompts: (1) ROI-level prompts encoding brain region identities and functions, (2) subject-level prompts incorporating demographic information, and (3) disease-level prompts capturing temporal disease progression. Compared to conventional methods that rigidly integrate demographic data, such as directly encoding numerical features, LLMs provide a flexible and semantically rich representation that can capture intricate dependencies, including various semantic and contextual relationships within the multi-modal data. By incorporating these prompts into graph learning, BrainPrompt leverages general structural, functional, and contextual insights of LLMs, enhancing both disease staging accuracy and interpretability. Our key contributions include:

- We introduce the first multi-level prompt enhancement framework for brain network analysis with fMRI data, leveraging non-imaging information and external knowledge from LLMs.
- We evaluate BrainPrompt on two fMRI datasets, showing its superiority over state-of-the-art methods in neurological condition identification. It also shows impressive adaptability by incorporating various backbones.
- We present a biomarker study that highlights the ability of our approach to extract interpretable pattern, which aligns with neuroscience knowledge.

2 Problem Definition

The task of neurological condition identification is to predict the stage of a given condition that the subject is currently experiencing. Given a labeled brain network dataset $\mathcal{D} = (\mathcal{G}, \mathcal{Y}) = \{(G, y_G)\}$, where y_G is the class label of a brain network $G \in \mathcal{G}$, the problem of graph classification is to learn a predictive function $f: \mathcal{G} \rightarrow \mathcal{Y}$ that maps input graphs to their labels. Herein, we use the brain networks provided by Xu et al. [29]. All preprocessed fMRI data are parcellated using the AAL atlas, resulting in 116 ROIs [22], i.e., the number of nodes for each subject $|\mathcal{V}_G| = 116$. We chose the AAL atlas due to its wide use and its ability to provide both anatomical and functional context. For each subject, we construct a connectivity matrix $\mathbf{C}$, where the nodes represent ROIs, and the edges encode the Pearson’s correlation between the blood-oxygen-level-dependent (BOLD) signals of each pair of ROIs. To represent the brain network as a graph $G = (\mathbf{X}, \mathbf{A})$, we define the feature matrix $\mathbf{X} = \mathbf{C}$, and the adjacency matrix $\mathbf{A}$ as a sparsified version of $\mathbf{C}$, retaining the top 20% correlations as connections, to avoid over-smoothing.

3 Proposed Method

In this section, we introduce our proposed method in detail, with the overall framework shown in Fig. 1. Specifically, BrainPrompt enhances GNNs through three levels of prompt enhancement: (1) an ROI-level prompt that encodes the general structure and function of each ROI to give prominence to its discriminative identity; (2) a subject-level prompt that incorporates the demographic information of the subject, which is highly complementary to embeddings based solely on fMRI images; and (3) a disease-level prompt that captures the sequential variations along the progression of degeneration, from healthy controls to the most deteriorated group. We employ the Llama-encoder-1.0B¹ of LLM2Vec [1] as the text encoder. Given the limited scale of fMRI datasets, we chose to freeze the LLM encoder during training to prevent overfitting.

3.1 Message-Passing with ROI Prompt Enhancement

In the brain atlas, each ROI is associated with a specific medical terminology, and the connectivity and activity of brain regions are critical for analyzing brain disorders. However, graph-structured brain networks are inherently permutation-invariant, meaning that nodes cannot inherently recognize their unique ROI identities [25]. Recent works [14, 27, 28] have proposed learnable identity embeddings for each ROI to address this limitation. These methods independently learn identity embeddings for each ROI, but they do not account for the general structural and functional relationships between different ROIs. To bridge this gap, we enhance message-passing neural networks with the ROI-level prompt.

¹ <https://huggingface.co/knowledgator/Llama-encoder-1.0B>

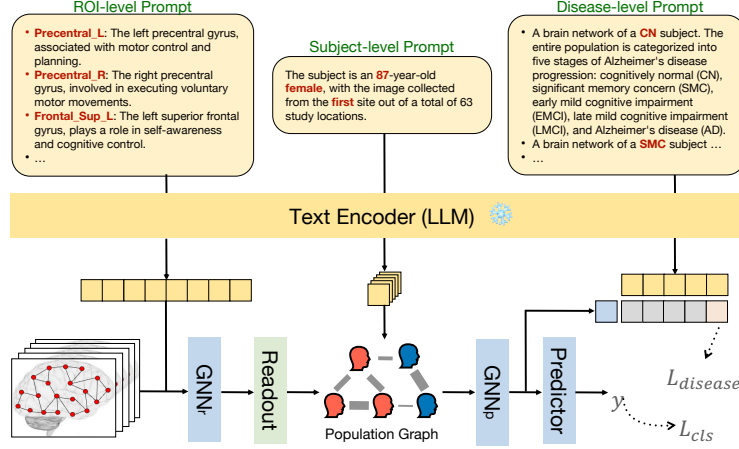


Fig. 1. The architecture of BrainPrompt, using Alzheimer’s Disease (AD) as an example. To extract more discriminative information, we do not utilize the entire representation of the used prompts for subsequent computations. Instead, we focus on the portion of tokens of each prompt that differ from other instances, which we define as informative tokens (e.g., age, sex, site, and labels, denoted by **red color**).

We first define the l -th layer of a GNN in message-passing scheme [30] as:

$$\mathbf{h}_v^{(l)} = \text{GNN}^{(l-1)}(\mathbf{h}_v^{(l-1)}) = \text{AGG}^{(l-1)}\left(\mathbf{h}_v^{(l-1)}, \text{MSG}^{(l-1)}\left(\left\{\mathbf{h}_u^{(l-1)}\right\}_{\forall u \in \mathcal{N}(v)}\right)\right), \quad (1)$$

where $\mathbf{h}_v^{(l)} \in \mathbb{R}^d$ denotes the node representation of a node $v \in \mathcal{V}_G$ at the l -th layer and d is the hidden dimension. $\text{AGG}(\cdot)$ and $\text{MSG}(\cdot)$ are arbitrary differentiable aggregate and message functions (e.g., a multilayer perceptron (MLP) as $\text{AGG}(\cdot)$ and a summation function as $\text{MSG}(\cdot)$). $\mathcal{N}(v)$ denotes the neighbors of node v defined by $\mathbf{A}$, and $\mathbf{h}_v^{(0)} = \mathbf{X}[v, :]$ representing the raw features of node v .

To incorporate ROI identity into graph representation learning, we generate prompts from the ROI labels (e.g., Angular_L, Caudate_R, etc.) and pass them through the text encoder. The l -th GNN layer with the ROI-level prompt is:

$$\mathbf{h}_v^{(l)} = \text{GNN}_r^{(l-1)}\left(\mathbf{h}_v^{(l-1)} + \text{Enc}(\mathbf{p}_v^{ROI})\right), \quad (2)$$

where $\mathbf{p}_v^{ROI}$ denotes the prompt of node v , and $\text{Enc}(\cdot)$ denotes the text encoder with a linear transformation to align the dimensionality of the embeddings from the pre-trained language model with the hidden dimension of the GNN layer. The ROI prompt $\mathbf{p}_v^{ROI}$ is generated by ChatGPT using the ROI terminology specific to the brain atlas. Specifically, the query we sent to ChatGPT is “Given the ROI labels for AAL116 atlas, generate a sentence to describe each of them by the given order: Precentral_L, Precentral_R, Frontal_Sup_L ...”. The output is as shown in Fig. 1, these ROI prompts are derived from the ROI labels to describe the general structural and functional features of each region. Domain experts review these prompts to ensure the clinical relevance and correctness.

3.2 Subject-Prompted Population Graph

In neurological condition identification, demographic information such as sex, age, and acquisition site provides crucial complementary insights beyond image-based embeddings. Prior works [10] encode demographic data as categorical features, which fail to capture the continuity and semantic richness inherent in such clinical data. Instead, we introduce a subject-level prompt $\mathbf{p}^{subject}$, formulated as: “The subject is a(n) [AGE]-year-old [SEX], with the image collected from the [SITE_NAME] site out of a total of [TOTAL_NUMBER_OF_SITES] study locations”. As in Fig. 1, this prompt encodes subject-specific attributes in natural language. To integrate subject-level prompts with imaging data, we construct a population graph $G^{pop} = (\mathbf{M}, \mathbf{A}')$ for each batch of subjects. The feature matrix $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_{bz}]^T \in \mathbb{R}^{bz \times d}$ is computed by aggregating node embeddings of each subject using: $\mathbf{m} = \text{READOUT}(\{\mathbf{h}_v\}_{v \in \mathcal{V}_G})$, where bz is the batch size and $\text{READOUT}(\cdot)$ is a sum pooling. The adjacency matrix $\mathbf{A}'$ is constructed based on both node embedding similarity and demographic similarity of the subjects:

$$\mathbf{S}_1(i, j) = \text{sim}(\mathbf{m}_i, \mathbf{m}_j), \mathbf{S}_2(i, j) = \text{sim}(\text{Enc}(\mathbf{p}_i^{subject}), \text{Enc}(\mathbf{p}_j^{subject})), \quad (3)$$

where $\text{sim}(\cdot, \cdot)$ denotes the cosine similarity between two vectors. The binarized adjacency matrix $\mathbf{A}'$ is obtained by thresholding the Hadamard products of $\mathbf{S}_1$ and $\mathbf{S}_2$, setting entries greater than a preset threshold th to 1 and others to 0.

A single-layer GNN named $\text{GNN}_p(\cdot)$, following the definition in Eq. (1), is applied to refine subject representations. Finally, for graph classification, we employ a two-layer MLP as predictor with a cross-entropy loss L_{cls} .

3.3 Temporal Constraint with Disease-Prompt

Inspired by CLIP [18], we further introduce a disease-level prompt to represent the label of each subject and incorporate a temporal constraint for the condition. Specifically, this is achieved by comparing the subject representation with text embeddings that correspond to textual descriptions of different stages of the condition. Let $\{\mathbf{p}_i^{disease}\}_{i=1}^k$ be the set of disease-level prompts, where k denotes the number of classes. Each prompt $\mathbf{p}_i^{disease}$ is derived from a text template of the form “A brain network of a(n) [CLASS] subject ...”, with the class token substituted by the specific class name. An example of a disease-level prompt is shown in Fig. 1. The disease-level prompts are encoded using $\text{Enc}(\cdot)$ as well, and another cross-entropy loss $L_{disease}$ is computed by comparing the subject representation with the embedding of these prompts. The overall loss of BrainPrompt is defined as $L = L_{cls} + \lambda * L_{disease}$, where λ is a trade-off hyperparameter.

4 Experiments

4.1 Datasets

We use two brain network datasets constructed by Xu et al. [29] for different disorders, which are ABIDE [4] for Autism Spectrum Disorder (ASD) and ADNI [5] for AD. Statistics of the brain network datasets are summarized in Table 1.

Table 1. Statistics and class information of brain network datasets used in this work.

Dataset	Condition	Sex (F/M)	Age (mean $\pm$ std)	Subject#	Class#
ABIDE	Autism Spectrum Disorder	152/873	16.5 $\pm$ 7.4	1025	2
ADNI	Alzheimer’s Disease	728/599	74.6 $\pm$ 7.9	1326	5

ABIDE The ABIDE initiative supports the research on ASD by aggregating functional brain imaging data from laboratories worldwide. ASD is characterized by stereotyped behaviors, including irritability, hyperactivity, depression, and anxiety. Subjects in the dataset are classified into two groups: Typical Control (TC) and individuals diagnosed with ASD.

ADNI The ADNI raw images used in this paper were obtained from the ADNI database (adni.loni.usc.edu). The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For up-to-date information, see www.adni-info.org. We include subjects from 5 different stages of AD, from cognitive normal (CN), significant memory concern (SMC), early MCI (EMCI), late MCI (LMCI) to AD.

4.2 Experimental Setup

We select various baselines, including (1) general-purpose GNN: **GCN** [11], (2) models tailored for image-only brain networks: **BrainNetCNN** [9], **BrainGNN** [14], **BNT** [8], **ContrastPool** [25], and (3) Demographic-enhanced GNN: **LG-GNN** [31], and (4) LLM-incorporated model: **ftSPL** [23].

The settings of our experiments mainly follow those in Dwivedi et al. [6]. We split each dataset into 8:1:1 for training, validation and test, respectively. We evaluate each model with the same random seed under 10-fold Cross-Validation (CV) and report the average accuracy. We use the early stopping criterion, i.e., we halve the learning rate when there is no further improvement on the validation loss during 25 epochs and stop the training once the learning rate is smaller than $1e-6$. We tuned the other hyperparameters on the validation set including the threshold th (from $\{0.4, 0.6, 0.8\}$), trade-off hyperparameters λ for the loss function (from $\{10.0, 1.0, 0.1\}$), the initial learning rate (from $\{1e-4, 5e-5, 1e-5\}$), and the weight decay (from $\{1e-2, 5e-3, 1e-3\}$) by grid search.

4.3 Main Results

The quantitative results on the ADNI and ABIDE datasets are presented in Table 2. We evaluate two variants of our proposed BrainPrompt by implementing GNN_r with GCN and BrainNetCNN, which are named as BrainPrompt-G and BrainPrompt-C respectively in the table. We can observe that the proposed BrainPrompt outperforms all seven baseline models across both datasets. The p-values of one-sided paired t-test comparing BrainPrompt with the best baselines on ADNI and ABIDE datasets are 0.0028 and 0.0021 respectively. Notably,

BrainPrompt enhances the accuracy of its GCN and BrainNetCNN backbones by up to 10.71% and 5.37%, respectively. These results underscore the effectiveness of our multi-level prompt enhancement approach for brain network analysis. It also demonstrates the adaptability of BrainPrompt on not only GNN-based methods but also CNN-based models.

Table 2. Results (average accuracy $\pm$ standard deviation) over 10-fold-CV. The best result is highlighted in **bold** while the runner-up is highlighted in underlined.

Model	ADNI	ABIDE					
	Accuracy	Accuracy	Precision	Recall	F1	ROC-AUC	
GCN	62.87 $\pm$ 2.92	60.10 $\pm$ 5.74	57.83 $\pm$ 5.87	60.03 $\pm$ 7.87	58.79 $\pm$ 6.32	60.08 $\pm$ 5.76	
BrainNetCNN	65.40 $\pm$ 2.22	64.58 $\pm$ 6.29	63.10 $\pm$ 6.82	62.09 $\pm$ 6.92	62.53 $\pm$ 6.53	64.46 $\pm$ 6.28	
BrainGNN	64.76 $\pm$ 3.06	61.39 $\pm$ 6.42	59.43 $\pm$ 6.96	61.19 $\pm$ 7.22	60.13 $\pm$ 6.28	61.37 $\pm$ 6.34	
BNT	64.59 $\pm$ 5.91	58.95 $\pm$ 4.84	58.62 $\pm$ 4.34	61.10 $\pm$ 13.46	59.13 $\pm$ 7.41	60.91 $\pm$ 4.58	
ContrastPool	66.17 $\pm$ 4.61	63.63 $\pm$ 3.26	57.65 $\pm$ 7.28	72.13 $\pm$ 6.06	62.78 $\pm$ 6.68	65.08 $\pm$ 3.32	
LG-GNN	67.28 $\pm$ 2.46	63.96 $\pm$ 4.22	63.03 $\pm$ 4.78	59.41 $\pm$ 7.94	60.90 $\pm$ 5.60	63.73 $\pm$ 4.29	
FTSPL	64.66 $\pm$ 2.79	63.49 $\pm$ 5.73	63.85 $\pm$ 6.10	<u>63.25</u> $\pm$ 5.72	62.98 $\pm$ 5.81	66.07 $\pm$ 5.84	
BrainPrompt-G	70.41 $\pm$ 2.12	<u>65.82</u> $\pm$ 4.82	<u>64.17</u> $\pm$ 5.48	62.92 $\pm$ 5.22	<u>63.43</u> $\pm$ 4.65	<u>66.32</u> $\pm$ 4.53	
BrainPrompt-C	<u>68.91</u> $\pm$ 3.10	66.20 $\pm$ 5.93	65.31 $\pm$ 6.55	62.07 $\pm$ 9.05	63.45 $\pm$ 7.26	67.00 $\pm$ 6.01	

4.4 Ablation Study

To assess the impact of each key component, we conduct experiments by disabling individual prompts while keeping all other settings unchanged. As shown in Table 3, introducing each prompt individually results in a measurable improvement. More interestingly, we observe strong synergistic effects when multiple components are combined. Specifically, integrating ROI-level and subject-level prompts leads to a significant accuracy boost of +3.93% (ADNI) and +4.85% (ABIDE) over the baseline. Similarly, combining ROI-level and disease-level prompts results in an improvement of +3.55% (ADNI) and +1.62% (ABIDE). These findings underscore that no single component alone is sufficient for optimal performance. Instead, their combined use enables BrainPrompt-G to capture general structural, functional, and contextual information, ultimately leading to more precise disease prediction.

4.5 Biomarker Detection

Despite the high accuracy achieved by our model, a critical concern is the interpretability of its decision-making process. In the context of brain biomarker detection, identifying salient ROIs associated with predictions as potential biomarkers is essential. To interpret BrainPrompt’s reasoning, we visualize the top 10 most salient ROIs using integrated gradients [20] for AD and ASD on the ADNI and ABIDE datasets, employing the Captum package [13]. As shown in Fig. 2, highlighted ROIs, such as the right inferior occipital gyrus, left inferior frontal

Table 3. Ablation study on important modules in BrainPrompt-G on ADNI and ABIDE datasets. The best result (accuracy) is highlighted in **bold**. We treat the first line as the GCN baseline, and Δ denotes the difference compared to the baseline.

ROI Subject Disease	ADNI	ABIDE	Δ_{ADNI}	Δ_{ABIDE}
	62.87 ± 2.92	60.10 ± 5.74	0.00	0.00
✓	65.19 ± 2.20	61.77 ± 5.55	+2.32	+1.67
✓	63.92 ± 4.29	62.30 ± 6.39	+1.05	+2.20
✓	61.95 ± 4.26	60.91 ± 5.36	-0.92	+0.81
✓	66.80 ± 2.83	64.95 ± 5.38	+3.93	+4.85
✓	66.42 ± 2.23	61.72 ± 2.92	+3.55	+1.62
✓	64.16 ± 4.34	61.33 ± 5.76	+1.29	+1.23
✓	70.41 ± 2.12	65.82 ± 4.82	+7.54	+5.72

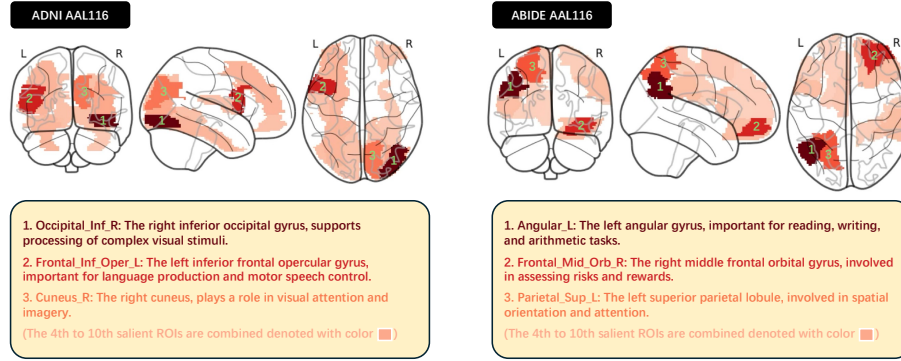


Fig. 2. Visualization of the top 10 salient ROIs according to the integrated gradients of BrainPrompt-G (the magnitude of saliency is represented by color intensity). We also show some of the corresponding ROI-level prompts.

opercular gyrus, and right cuneus in the ADNI dataset, suggest potential AD-specific neural mechanisms. These regions have been previously identified in Alzheimer’s studies [32, 7] as key areas linked to cognitive functions. For ASD, BrainPrompt highlights ROIs corresponding to the left angular gyrus, right middle frontal orbital gyrus and left superior parietal lobule in the ABIDE dataset, which align with domain knowledge from prior Autism research [2, 12].

5 Conclusion

In this paper, we propose BrainPrompt, a novel framework that integrates multi-level prompt enhancement for brain network analysis in neurological condition identification. By incorporating ROI-level, subject-level, and disease-level prompts, BrainPrompt effectively captures both the general structural and functional features of brain networks as well as critical contextual information about the subject and disease progression. Our experimental results on the ADNI and ABIDE datasets demonstrate that BrainPrompt consistently outperforms state-

of-the-art baselines, achieving significant improvements in classification accuracy. These findings highlight the potential of leveraging non-imaging information and external knowledge from LLMs to improve the interpretability and accuracy of brain network analysis, offering valuable insights for the diagnosis and staging of neurological diseases. Future work will explore extending this approach to additional tasks and other neuroimaging modalities to further validate the generalizability of BrainPrompt.

Acknowledgments. This research/project is supported by the Ministry of Education, Singapore under its Academic Research Fund Tier 1 (RG16/24). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Ministry of Education, Singapore.

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

References

1. BehnamGhader, P., Adlakha, V., Mosbach, M., Bahdanau, D., Chapados, N., Reddy, S.: LLM2Vec: Large language models are secretly powerful text encoders. In: First Conference on Language Modeling (2024)
2. Chen, S., Xing, Y., Kang, J.: Latent and abnormal functional connectivity circuits in autism spectrum disorder. *Frontiers in neuroscience* **11**, 125 (2017)
3. Cho, H., Sim, J., Wu, G., Kim, W.H.: Neurodegenerative brain network classification via adaptive diffusion with temporal regularization. In: Forty-first International Conference On Machine Learning (2024)
4. Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., Khundrakpam, B.S., Lewis, J.D., Li, Q., Milham, M., et al.: The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics* **7**, 27 (2013)
5. Dadi, K., Rahim, M., Abraham, A., Chyzyk, D., Milham, M., Thirion, B., Varoquaux, G., Initiative, A.D.N., et al.: Benchmarking functional connectome-based predictive models for resting-state fmri. *NeuroImage* **192**, 115–134 (2019)
6. Dwivedi, V.P., Joshi, C.K., Luu, A.T., Laurent, T., Bengio, Y., Bresson, X.: Benchmarking graph neural networks. *arXiv preprint arXiv:2003.00982* (2020)
7. He, Y., Wang, L., Zang, Y., Tian, L., Zhang, X., Li, K., Jiang, T.: Regional coherence changes in the early stages of alzheimer’s disease: a combined structural and resting-state functional mri study. *Neuroimage* **35**(2), 488–500 (2007)
8. Kan, X., Dai, W., Cui, H., Zhang, Z., Guo, Y., Yang, C.: Brain network transformer. *Advances in Neural Information Processing Systems* **35**, 25586–25599 (2022)
9. Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G.: Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage* **146**, 1038–1049 (2017)
10. Kazi, A., Shekarforoush, S., Kortuem, K., Albarqouni, S., Navab, N., et al.: Self-attention equipped graph convolutions for disease prediction. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 1896–1899. IEEE (2019)

11. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017)
12. Kohls, G., Schulte-Rüther, M., Nehr Korn, B., Müller, K., Fink, G.R., Kamp-Becker, I., Herpertz-Dahlmann, B., Schultz, R.T., Konrad, K.: Reward system dysfunction in autism spectrum disorders. *Social cognitive and affective neuroscience* **8**(5), 565–572 (2013)
13. Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., Melnikov, A., Kliushkina, N., Araya, C., Yan, S., et al.: Captum: A unified and generic model interpretability library for pytorch. arXiv preprint arXiv:2009.07896 (2020)
14. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L.H., Ventola, P., Duncan, J.S.: Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical Image Analysis* **74**, 102233 (2021)
15. Mielke, M.M., Vemuri, P., Rocca, W.A.: Clinical epidemiology of alzheimer’s disease: assessing sex and gender differences. *Clinical epidemiology* pp. 37–48 (2014)
16. Mistur, R., Mosconi, L., De Santi, S., Guzman, M., Li, Y., Tsui, W., de Leon, M.J.: Current challenges for the early detection of alzheimer’s disease: brain imaging and csf studies. *Journal of clinical neurology* **5**(4), 153–166 (2009)
17. Perochon, S., Di Martino, J.M., Carpenter, K.L., Compton, S., Davis, N., Eichner, B., Espinosa, S., Franz, L., Krishnappa Babu, P.R., Sapiro, G., et al.: Early detection of autism using digital behavioral phenotyping. *Nature Medicine* **29**(10), 2489–2497 (2023)
18. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
19. Song, X., Frangi, A., Xiao, X., Cao, J., Wang, T., Lei, B.: Integrating similarity awareness and adaptive calibration in graph convolution network to predict disease. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23. pp. 124–133. Springer (2020)
20. Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. In: International conference on machine learning. pp. 3319–3328. PMLR (2017)
21. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al.: Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023)
22. 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)
23. Wang, P., Zhang, H., He, Z., Peng, Z., Yuan, Y.: ftspl: Enhancing brain analysis with fmri-text synergistic prompt learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 564–574. Springer (2024)
24. Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G., Morales, F., Evans, A.C.: A general statistical analysis for fmri data. *Neuroimage* **15**(1), 1–15 (2002)
25. Xu, J., Bian, Q., Li, X., Zhang, A., Ke, Y., Qiao, M., Zhang, W., Sim, W.K.J., Gulyás, B.: Contrastive graph pooling for explainable classification of brain networks. *IEEE Transactions on Medical Imaging* (2024)

26. Xu, J., Chen, Y., Dong, X., Lan, M., HUANG, T., Bian, Q., Cheng, J., Ke, Y.: BrainOOD: Out-of-distribution generalizable brain network analysis. In: The Thirteenth International Conference on Learning Representations (2025), <https://openreview.net/forum?id=3xqqYOKILp>
27. Xu, J., He, K., Lan, M., Bian, Q., Li, W., Li, T., Ke, Y., Qiao, M.: Contrasformer: A brain network contrastive transformer for neurodegenerative condition identification. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. pp. 2671–2681 (2024)
28. Xu, J., Lan, M., Dong, X., He, K., Zhang, W., Bian, Q., Ke, Y.: Multi-atlas brain network classification through consistency distillation and complementary information fusion. arXiv preprint arXiv:2410.08228 (2024)
29. Xu, J., Yang, Y., Huang, D.T.J., Gururajapathy, S.S., Ke, Y., Qiao, M., Wang, A., Kumar, H., McGeown, J., Kwon, E.: Data-driven network neuroscience: On data collection and benchmark. In: Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track (2023)
30. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: International Conference on Learning Representations (2019)
31. Zhang, H., Song, R., Wang, L., Zhang, L., Wang, D., Wang, C., Zhang, W.: Classification of brain disorders in rs-fMRI via local-to-global graph neural networks. IEEE Transactions on Medical Imaging (2022)
32. Zhang, T., Davatzikos, C.: Optimally-discriminative voxel-based morphometry significantly increases the ability to detect group differences in schizophrenia, mild cognitive impairment, and Alzheimer's disease. Neuroimage **79**, 94–110 (2013)