G - LaD: Graph-Language alignment for few-shot Diagnosis from fMRI

Abhishek Gupta[1], **Vipul Kumar Singh**[2], **Jyotismita Barman**[2], **Sandeep Kumar**[1,2,3], **Anish Arora**[4]
[1]Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, India
[2]Department of Electrical Engineering, Indian Institute of Technology Delhi, India
[3]Bharti School of Telecommunication Technology and Management, Indian Institute of Technology Delhi, India
[4]Department of Computer Science and Engineering, The Ohio State University, USA
Abhishek.Gupta@scai.iitd.ac.in, Vipul.Kumar.Singh@ee.iitd.ac.in, Jyotismita.Barman@ee.iitd.ac.in, ksandeep@iitd.ac.in, anish@cse.ohio-state.edu

Abstract

Decoding human cognitive states from neural activity is a core challenge in artificial intelligence and computational neuroscience. Functional Magnetic Resonance Imaging (fMRI) captures high-dimensional spatiotemporal patterns of brain activity, yet characterizing cognitive states based on modeling the complex, dynamic dependencies among distributed regions remains difficult. While Graph Neural Networks (GNNs) to represent the brain as a structured graph has advanced functional connectivity (FC) analysis, they suffer from limited generalization, reliance on large labeled datasets, and poor transferability across neuro-imaging tasks. We introduce **G**raph-**L**anguage **a**lignment for **D**iagnosis (G-LaD), that integrates graph representation learning with Large Language Models (LLMs) for data-efficient brain graph classification. G-LaD first pretrains a graph encoder, built upon Graph Isomorphism Network layers using a reconstruction-driven Denoising Autoencoder, to capture structural and topological invariants. In the second stage, distribution-level alignment between graph and language representations is achieved via a Sinkhorn-divergence objective, enabling smooth and transferable cross-modal mapping. Finally, a Chain-of-Thought prompting mechanism guides the LLM to perform reasoning-driven predictions. Empirical evaluations on the ABIDE dataset demonstrate superior few-shot generalization and robust performance of G-LAD in neurodegenerative disorder classification.

1 Introduction

Understanding and decoding human cognitive states from neural activity remains a central challenge in artificial intelligence and computational neuroscience (Mitchell et al. 2004). Functional Magnetic Resonance Imaging (fMRI) provides a powerful, noninvasive window into brain dynamics by measuring blood-oxygen-level-dependent (BOLD) signals (Logothetis and Wandell 2004). These signals form a high-dimensional spatiotemporal map that serves as an established proxy for neural activation patterns. The fMRI signal, however, is inherently noisy and high dimensional, reflecting complex interactions among spatially distributed brain regions. Traditional machine learning models, which rely on handcrafted features or linear assumptions, struggle to capture these nonlinear and dynamic dependencies (Abraham et al. 2017; Lee et al. 2014).

Deep learning has shown growing capability for fMRI analysis (Wen et al. 2018). Convolutional Neural Networks (CNNs) can extract local spatial features but often impose grid-like assumptions that disregard the intrinsic functional organization of the brain. Graph Neural Networks (GNNs) offer a more biologically grounded representation by modeling the brain as a graph in which regions of interest (ROIs) correspond to nodes and functional connections define the edges (Bessadok, Mahjoub, and Rekik 2022). Despite their success, graph-based approaches face persistent challenges. Their generalization is often limited by class imbalance, dependence on large labeled datasets, and poor transferability across diagnostic or cognitive tasks. These issues underscore the need for more adaptable and data-efficient graph learning frameworks capable of capturing the complexity and variability of brain functional connectivity. To mitigate the reliance on labeled data, self-supervised graph learning methods have been introduced (Peng et al. 2022). Yet, these techniques typically require extensive task-specific fine-tuning to perform well on downstream objectives, restricting their utility in few-shot and zero-shot settings.

In light of these limitations, Large Language Models (LLMs) have recently emerged as powerful universal learners, showing strong performance in few-shot and zero-shot graph reasoning tasks, particularly on knowledge and text-attributed graphs (Chen et al. 2024b; Jin et al. 2024). However, translating graph structure into natural language often leads to suboptimal representations, as the contextual dependencies among neighboring nodes are not effectively preserved(Huang et al. 2023). Some studies have explored using LLMs to generate enriched node embeddings that are subsequently processed by GNNs, but these approaches remain constrained by the predictive limitations of traditional GNN architectures (Chen et al. 2024a; Tang et al. 2024).

In this paper, we leverage LLMs as graph predictors and demonstrate that LLMs are capable of reasoning directly over neuro-imaging graphs. In doing so, we overcome the SOTA limitations wherein LLM performance as graph predictors has been inconsistent. The central challenge lies in learning transferable graph representations that generalize across heterogeneous datasets and tasks when interacted with LLMs. Moreover, the ability of language models to reason directly over neuro-imaging graphs remains largely unexplored.

The core idea underlying language model–based predictor

Method	Self-supervised Representation	Few-shot / Label Efficient Focus	Graph–Language Alignment	Balanced Diagnostic Objective
BrainNetCNN	✗	✗	✗	✗
BrainGNN	✗	✗	✗	✗
GCDA	✓	✓	✗	✗
ALTER	✗	✗	✗	✗
G-LaD (Ours)	✓	✓	✓	✓

Table 1: Comparison of G-LaD with representative brain network models. A ✓ indicates the presence of a desirable property, while ✗ indicates its absence.

frameworks is to pre-train a functional Encoder and align its latent representations with the token embeddings of a frozen LLM. However, existing approaches typically perform instance-level alignment, which may be insufficient for capturing global structural semantics across diverse graph tasks.

To address these limitations, we introduce **G**raph-**L**anguage **a**lignment for **D**iagnosis (G-LaD), a novel two-stage framework for brain functional connectivity analysis. In the first stage, we pre-train a graph encoder built upon Graph Isomorphism Network (GIN) (Xu et al. 2018) layers using a denoising Autoencoder architecture (Vincent et al. 2008) to learn robust topological representations. In the second stage, rather than relying on instance-level alignment, we employ distribution-level alignment through a Sinkhorn-divergence–based training objective (Oneto et al. 2020), promoting smoother and more transferable representation matching. Finally, we incorporate Chain-of-Thought (CoT) instructions (Wei et al. 2022) to guide the LLM reasoning in producing the final predictions.

To the best of our knowledge, this study is the first to introduce a CoT–driven LLM predictor for brain graph classification. Comprehensive experiments validate the effectiveness of the proposed framework, highlighting its capability for robust few-shot generalization and accurate prediction of neuro-degenerative disorder. The contributions of this study are summarized as follows:

- We propose G-LaD, the first framework that integrates graph-based brain representations with large language models through a distribution-level alignment mechanism.
- We develop a trainable projection module trained with a Sinkhorn-divergence–based objective, establishing a bridge between the graph embedding space and the LLM token embedding space for structure–semantic alignment.
- We introduce an auxiliary balancing loss that explicitly mitigates class imbalance in neuroimaging datasets.
- Extensive experiments on the ABIDE benchmark dataset (Di Martino et al. 2014) demonstrate the effectiveness of the proposed framework.

2 Problem Description and Related Work

2.1 Problem Description

Let each subject be denoted as S_i. The corresponding BOLD signal is represented as $\mathbf{B}_i \in \mathbb{R}^{n \times T_i}$, where n denotes the number of Regions of Interest (ROIs) and T_i represents the temporal length of the fMRI scan for subject i. Each row of $\mathbf{B}_i$ encodes the time series of the BOLD response for a specific ROI. From the BOLD signals, we derive a functional connectivity (FC) matrix $\mathbf{A}_i \in \mathbb{R}^{n \times n}$, where each element a_{pq} quantifies the statistical dependency (e.g., Pearson correlation, partial correlation, or mutual information) between the p^{th} and q^{th} ROIs. Each ROI is further associated with a d-dimensional feature vector, forming the node-feature matrix $\mathbf{X}_i \in \mathbb{R}^{n \times d}$.

Hence, each subject S_i can be expressed as a weighted graph $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i, \mathbf{X}_i, \mathbf{A}_i)$, where $\mathcal{V}_i$ and $\mathcal{E}_i$ denote the sets of nodes (ROIs) and edges (functional connections), respectively. The objective is to learn a parameterized mapping

$$f_{\boldsymbol{\Theta}} : (\mathbf{X}_i, \mathbf{A}_i) \mapsto \mathbf{y}_i,$$

where $f_{\boldsymbol{\Theta}}$ represents a classifier, parameterized by $\boldsymbol{\Theta}$, that predicts the diagnostic or cognitive state label $\mathbf{y}_i$ of subject S_i. The goal is to optimize $f_{\boldsymbol{\Theta}}$ such that it generalizes across subjects and tasks under limited supervision, enabling robust few-shot prediction of neurological disorders. This problem setup has inspired diverse methodological advances in neuro-imaging based disorder prediction, as discussed in the following related works.

2.2 Related Work

Brain Functional Analysis. Several studies have explored graph learning paradigms for brain network analysis, addressing tasks such as neurological disorder diagnosis and biological sex prediction. Informative features are extracted from the fMRI data using techniques such as clustering-based methods and decomposition-based frameworks (Wee et al. 2012; Tang et al. 2016). Then, these features are used for disease prediction through classical machine-learning models, including ridge classifiers, logistic regression, and support vector machines (Abraham et al. 2017; Lee et al. 2014; Morra et al. 2009). These methods often neglect the essential topological properties of brain networks, which characterize the complex interactions among distinct brain regions.

In recent years, GNNs have emerged as powerful architectures capable of exploiting the inherent structure of brain connectivity data (Bessadok, Mahjoub, and Rekik 2022). BrainGNN (Li et al. 2021) develops ROI-aware graph convolutional layers that leverages the topological and functional information of fMRI. (Kan et al. 2022a) introduced a dynamically learnable brain network that optimizes connectivity patterns during training. GCDA (Wang et al. 2025) enables self-supervised brain graph representation learning by leveraging a diffusion-based augmentation method.

Recently, there has been a growing interest in applying transformer architectures to brain network analysis in order to capture the long range dependencies. For example, BrainNetTF (Kan et al. 2022b) leverages connection profiles as effective and computationally efficient positional encodings, while ALTER (Yu et al. 2024) employs adaptive random-walk–based positional strategies to capture long-range dependencies across brain regions.

A small set of work has explored the use of LLMs for brain network analysis, and their role has been largely limited to incorporating auxiliary textual information (Wang et al. 2024; Xu et al. 2025). These approaches still rely on

GNNs as the primary classifiers and thus inherit their limited generalization capability.

Language Models for Graph Tasks. Recent efforts have explored coupling GNNs with LLMs by designing dual-encoder frameworks and aligning graph and text representations at the instance level, typically via contrastive learning or EM-based training. Representative models such as MoMu (Su et al. 2022), MoleculeSTM (Liu et al. 2023), and ConGraT (Brannon et al. 2023) build domain-specific encoders tailored to tasks like molecular analysis or social network mining, and rely on sample-wise graph–text pairs for supervision. Extensions including G2P2 (Wen and Fang 2023, 2024), GRENADE (Li, Ding, and Lee 2023), THLM (Zou et al. 2023), and GLEM (Zhao et al. 2022) enhance this paradigm through refined alignment objectives or iterative pseudo-labeling, but still operate within application-specific pipelines and constrained modality-matching protocols.

Despite recent progress, the application of LLMs to achieve few-shot generalization in brain functional connectivity analysis has not yet been investigated.

3 G-LAD Framework

In this section, we describe the proposed two-stage G-LaD framework for brain disease classification that integrates the learned fMRI graph representation with the token embedding space of a frozen LLM. The proposed framework is structured into two sequential stages, as illustrated in Figure 1. First, we construct a brain functional graph $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i, \mathbf{X}_i, \mathbf{A}_i)$ from the preprocessed fMRI region-aware time series data, and feed it to the first stage of the framework that uses Self-Supervised Graph Representation Learning to translate $\mathcal{G}_i$ into a compact, topology-aware embedding. In its second stage, G-LaD integrates the embedding with a frozen LLM for prompt-driven, few-shot classification, using distribution matching by training a Graph Adapter.

3.1 Functional Connectivity Estimation

Estimating functional connectivity (FC) from the BOLD signal is a critical step for constructing brain graphs suitable for representation learning with GNNs. For each subject S_i, we compute an ROI-wise similarity matrix $\mathbf{C}_i \in \mathbb{R}^{n\times n}$, defined as

$$\mathbf{C}_i = \begin{bmatrix} \mathrm{sim}(\mathbf{b}_i^{(1)}, \mathbf{b}_i^{(1)}) & \cdots & \mathrm{sim}(\mathbf{b}_i^{(1)}, \mathbf{b}_i^{(n)}) \\ \vdots & \ddots & \\ \mathrm{sim}(\mathbf{b}_i^{(n)}, \mathbf{b}_i^{(1)}) & \cdots & \mathrm{sim}(\mathbf{b}_i^{(n)}, \mathbf{b}_i^{(n)}) \end{bmatrix}, \quad (1)$$

where $\mathbf{b}_i^{(k)} = \mathbf{B}_i[k, :]$ denotes the BOLD time series of the k^{th} ROI for subject i, and $\mathrm{sim}(\cdot, \cdot)$ is a similarity function that captures pairwise temporal dependencies between ROIs.

Empirical studies suggest that distinct similarity metrics capture complementary aspects of neural connectivity. Among these, the Pearson correlation coefficient remains the most widely used measure for FC estimation, defined as

$$\rho(\mathbf{b}_i^{(p)}, \mathbf{b}_i^{(q)}) = \frac{\mathbb{E}\left[(\mathbf{b}_i^{(p)} - \mu_i^{(p)})(\mathbf{b}_i^{(q)} - \mu_i^{(q)})\right]}{\sigma_{\mathbf{b}_i^{(p)}}\, \sigma_{\mathbf{b}_i^{(q)}}}, \quad (2)$$

where $\mu_i^{(p)}$ and $\sigma_{\mathbf{b}_i^{(p)}}$ denote the mean and standard deviation of the BOLD signal for ROI p. Finally, a non-negative threshold τ is applied to the similarity matrix to construct the binary or weighted adjacency matrix $\mathbf{A}_i = \mathbb{I}\left[\mathbf{C}_i > \tau\right]$, which represents the final functional connectivity graph for subject S_i.

Node attributes for each region of interest (RoI) are obtained from partial correlation coefficients, computed via the inverse covariance matrix of the BOLD signals, and represented as $\mathbf{X}_i$. In this study, we employed Pearson correlation to construct the brain connectivity graphs, while partial correlation coefficients were used as graph features.

3.2 Self-Supervised Graph Representation Learning via Denoising Autoencoder

To extract robust and invariant representations from brain graphs, we employ a self-supervised pretraining stage formulated as a Denoising Auto-Encoder (DAE). The objective is to learn an encoder capable of mapping each subject-level brain graph into a latent representation that preserves intrinsic structural connectivity while being resilient to noise or partial corruption.

The DAE consists of two jointly optimized components: an encoder $\mathcal{F}$ and a decoder $\mathcal{D}$. The encoder $\mathcal{F}$ is parameterized using Graph Isomorphism Network layers to ensure expressive structural discrimination. For each subject S_i, a stochastic corruption is applied to the adjacency matrix, producing a noisy version $\mathbf{A}_i'$. The perturbed graph is denoted as $\mathcal{G}_i' = (\mathcal{V}_i, \mathcal{E}_i', \mathbf{X}_i, \mathbf{A}_i')$, and its latent representation is computed as $\mathbf{Z}_i = \mathcal{F}(\mathcal{G}_i') \in \mathbb{R}^{n\times f}$.

The decoder $\mathcal{D}$ reconstructs the original connectivity pattern by predicting edge probabilities between node embeddings through a bilinear interaction, $\hat{\mathbf{A}}_i = \sigma(\mathbf{Z}_i \mathbf{Z}_i^\top)$, where $\sigma(\cdot)$ denotes the logistic sigmoid function.

Both modules are optimized jointly using a binary cross-entropy reconstruction loss:

$$\mathcal{L}_{\text{rec}} = -\sum_{u,v\in\mathcal{V}_i} \Big[\mathbf{A}_i[u,v] \log \hat{\mathbf{A}}_i[u,v] + (1-\mathbf{A}_i[u,v]) \log(1-\hat{\mathbf{A}}_i[u,v])\Big]. \quad (3)$$

Through this denoising objective, the encoder learns topology-preserving representations that are invariant to stochastic perturbations and capture the latent structure of brain connectivity graphs, providing a strong initialization for subsequent alignment and prediction stages.

Subject-Level Representation. After the denoising pretraining stage, the encoder $\mathcal{F}$ is employed as a graph feature extractor to obtain subject-level representations. For each subject graph $\mathcal{G}_i$, the encoder produces node-level embeddings. To aggregate node-level information into a single subject-level embedding, we employ a permutation-invariant READOUT function. In this work, we use the mean pooling operation. The resulting vector $\mathbf{g}_i \in \mathbb{R}^f$ encodes the global topological and functional characteristics of the subject's brain network within a compact latent space.

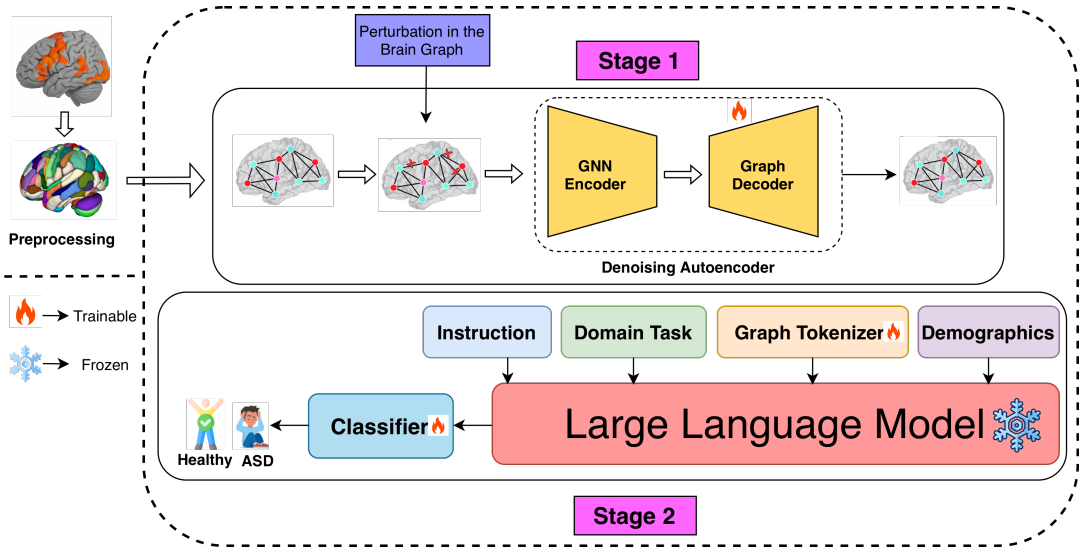

Figure 1: **Overall workflow of the G-LaD framework.** Stage 1 learns robust brain-graph embeddings via a self-supervised denoising graph autoencoder trained on functional connectivity graphs derived from fMRI. Stage 2 injects the learned embeddings, along with demographic features and task instructions, into a frozen Large Language Model for multimodal reasoning, while only a lightweight adapter and classifier are trained.

3.3 Multimodal Graph-Language Alignment

In the second stage of the G-LaD framework, a pretrained and frozen LLM is employed to process the subject-level representations, facilitating the extraction of disease-specific patterns for diagnostic prediction. This stage implements a *project-inject-reason-classify* architecture, where a lightweight, trainable adapter is optimized using a novel composite loss function, enabling the LLM to perform few-shot and CoT reasoning by directly integrating structural graph information with textual patient metadata.

To bridge the modality gap between the numerical graph-level representation $\mathbf{g}_i \in \mathbb{R}^f$ obtained from Stage 1 and the semantic embedding space of the frozen LLM, we introduce a learnable projector $\mathcal{P} : \mathbb{R}^f \to \mathbb{R}^{d_t}$, which maps graph embeddings into the token-embedding space of the language model.

Architecture and Training Objective. The second stage classifier is composed of three core components: (i) a trainable Graph Adapter $\mathcal{P}$ that projects the learned graph embeddings into the token embedding space of the language model, (ii) a frozen LLM backbone f_{LLM} that performs multimodal contextual reasoning, and (iii) a lightweight Classification Head ψ that maps the fused representation of the LLM to disease predictions. The adapter $\mathcal{P}$ and the head ψ are the only trainable modules in this stage, and all learning is driven by a unified composite loss that jointly optimizes: modality alignment, clinical discriminability, and class balance.

Graph–Language Adapter Network. The subject-level representation produced by the graph encoder resides in a numerical latent space that is inherently misaligned with the semantic token embedding manifold of the LLM. To enable effective cross-modal interaction, we introduce a learnable projection network $\mathcal{P}$ that maps graph-derived embeddings into the token embedding space of the frozen LLM.

Formally, for each subject representation $\mathbf{g}_i \in \mathbb{R}^f$, the adapter produces a projected embedding

$$\mathbf{g}_i^{\text{proj}} = \mathcal{P}(\mathbf{g}_i), \quad \mathbf{g}_i^{\text{proj}} \in \mathbb{R}^{d_{\text{model}}}, \tag{4}$$

where d_{model} denotes the dimensionality of the LLM token embedding space.

The adapter $\mathcal{P}$ is instantiated as a lightweight multi-layer perceptron and serves two central roles: (i) *modality bridging*, by transforming graph-derived embeddings into a representation compatible with the semantic token space of the language model, and (ii) *dimensionality alignment*, by mapping the graph embedding dimension to the token embedding dimensionality required by the pretrained LLM.

In particular, we employ a Sinkhorn-divergence based modality alignment loss which minimizes the discrepency between the distribution of the projected graph representations and the token embeddings. Let $\mathcal{Q}$ and $\mathcal{U}$ denote the distributions of the projected graph embeddings and a batch of LLM token embeddings, respectively. The alignment objective is defined as

$$\mathcal{L}_{\text{Sinkhorn}} = \mathbb{S}_\epsilon(\mathcal{Q}, \mathcal{U}), \tag{5}$$

where $\mathbb{S}(\cdot, \cdot)$ denotes the Sinkhorn-divergence.

Definition 1 (Sinkhorn Divergence) *Let $\mathcal{Q}$ and $\mathcal{U}$ be two probability distributions defined on a metric space $\mathcal{W}$ with cost function $c : \mathcal{W} \times \mathcal{W} \to \mathbb{R}_+$. The entropically regularized optimal transport objective is*

$$\text{OT}_\epsilon(\mathcal{Q}, \mathcal{U}) = \inf_{\pi \in \Pi(\mathcal{Q}, \mathcal{U})} \left[\int_{\mathcal{W} \times \mathcal{W}} c(x, y)\, d\pi(x, y) + \epsilon\, \mathcal{H}(\pi) \right], \tag{6}$$

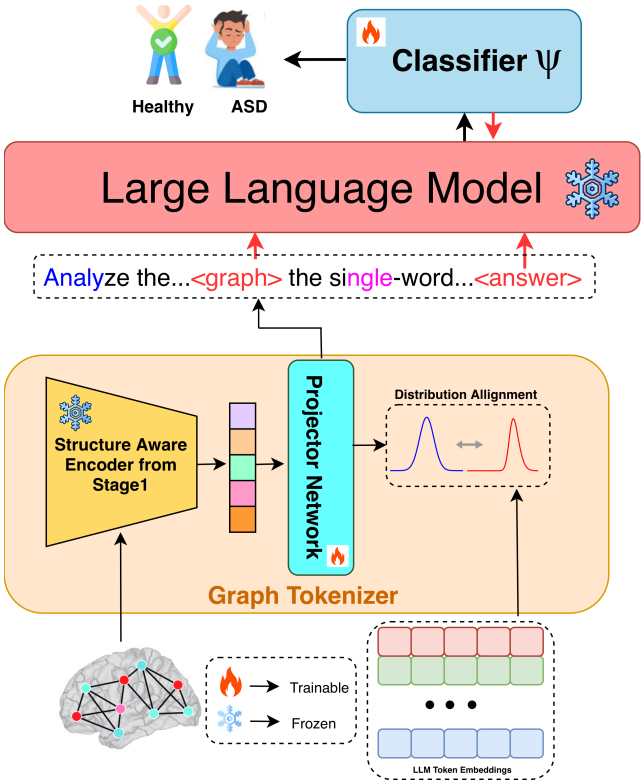

Figure 2: **Stage-2 multimodal alignment and reasoning pipeline of G-LaD.** The graph embedding from Stage-1 is projected into the LLM token space and injected into a structured prompt along with clinical information. The frozen LLM performs multimodal reasoning, while only the projection module $\mathcal{P}$ and classifier ψ are trained, enabling effective graph-to-language alignment for ASD classification.

where $\epsilon > 0$ is the entropic regularization coefficient and $\Pi(\mathcal{Q},\mathcal{U})$ denotes the set of joint couplings with marginals $\mathcal{Q}$ and $\mathcal{U}$. The entropy term $\mathcal{H}(\pi)$ is given by

$$\mathcal{H}(\pi) = \int_{\mathcal{W}\times\mathcal{W}} \log\left(\frac{d\pi(x,y)}{d\mathcal{Q}(x)\,d\mathcal{U}(y)}\right) d\pi(x,y). \quad (7)$$

The Sinkhorn divergence between $\mathcal{Q}$ and $\mathcal{U}$ is then defined as

$$\mathbb{S}_\epsilon(\mathcal{Q},\mathcal{U}) = \mathrm{OT}_\epsilon(\mathcal{Q},\mathcal{U}) - \frac{1}{2}\mathrm{OT}_\epsilon(\mathcal{Q},\mathcal{Q}) - \frac{1}{2}\mathrm{OT}_\epsilon(\mathcal{U},\mathcal{U}). \quad (8)$$

By minimizing this differentiable Sinkhorn-divergence $\mathcal{L}_{\text{Sinkhorn}}$, the projection network $\mathcal{P}$ learns to align the distribution of the graph embeddings with that of the LLM token space, thereby enabling effective cross-modal correspondence without updating the LLM parameters.

Multimodal Injection and Reasoning. Once the graph representation g_i^{proj} is projected into the token embedding space of langauge model, it is injected into a textual reasoning prompt to leverage the knowledge of LLM. Each training instance is formatted as a CoT instruction, which combines structured patient metadata with reasoning cues as shown in Figure 3.

The prompt is tokenized to obtain a sequence of IDs $J = [j_1, \ldots, j_T]$, which is further converted into token embeddings $E = Embed(J) \in \mathbb{R}^{T\times d_{\text{model}}}$. The position j_{graph} corresponds to the <|graph|> token, which serves as the injection point for the projected embedding g_i^{proj}, which means:

$$E[j_{graph}] \leftarrow g_i^{\text{proj}}, \quad (9)$$

producing a multimodal prompt sequence E_{mod}. This modified sequence is passed through the frozen LLM f_{LLM}:

$$H = f_{\text{LLM}}(E_{\text{mod}}), \quad \text{where } H = [h_1, h_2, \ldots, h_T] \quad (10)$$

Here, the LLM integrates g_i^{proj} with the natural language prompt, enabling its internal attention layers to process the injected graph representation as if it were a native semantic token.

Answer Extraction and Discriminative Alignment. The LLM output sequence H contains contextualized hidden states for all generated tokens, but the prediction of the model is fixed to the position of <|answer|> token. The hidden vector at that token position, $h_{\text{ans}} = H[i_{ans}]$, summarizes the multimodal reasoning of the LLM, combining textual context, demographic information and the functional connectivity-derived knowledge. This fused representation is then passed

Stage 2: Chain-of-Thought (CoT) Prompt

Instruction. Analyze the provided brain graph token and subject demographics for a given subject to classify Autism Spectrum Disorder (ASD) vs. Healthy Control (HC). First, provide a step-by-step analysis based on the graph token and subject demographics. Finally, conclude with the single-word classification: ASD or Control.

Demographics (Example). 24y Male, Right-Handed, Site NYU, FIQ 110, VIQ 112, PIQ 108, FD 0.120mm.

Graph Token. <|graph|>

Analysis and Final Answer. <|answer|>

Figure 3: **The complete Chain-of-Thought (CoT) prompt used in Stage 2.** This text is fed to the LLM, combining (1) a CoT instruction, (2) formatted patient demographics, and (3) two special tokens. The <|graph|> token is a placeholder where the projected graph embedding ($\mathbf{g}^{\text{proj}} = \mathcal{P}(g)$) is injected. The final hidden state at the <|answer|> token's position (h_{ans}) is extracted and passed to the classification head (ψ).

Data Split	Total Subjects	HC	ASD
Train	653	301	352
Validation	82	38	44
Test	82	38	44

Table 2: **Dataset partition for ABIDE.** Subject distribution across training, validation, and test sets, illustrating proportional representation of Healthy Controls (HC) and individuals with Autism Spectrum Disorder (ASD).

to the Classification Head ψ, which is a trainable MLP represented as $L = \psi(h_{ans})$, where $L \in \mathbb{R}^C$ produces logits over C disease categories. In this work, we focus on the standard binary classification paradigm for neurological disorder prediction commonly adopted in the neuroimaging literature. To ensure meaningful training, ψ and $\mathcal{P}$ are optimized not by standard cross entropy loss but via a ranking-based AUC loss $\mathcal{L}_{AUC}$, which directly maximizes diagnostic discrimination,

$$\mathcal{L}_{AUC} = \mathbb{E}_{(s_{pos}, s_{neg})}\Big[\text{softplus}(-(s_{pos} - s_{neg}))\Big] \quad (11)$$

where softplus(x) = $\ln(1 + \exp^x)$, s_{pos} and s_{neg} denote the logits corresponding to positive and negative cases within a batch. This explicitly increases the margin between positive and negative predictions by directly optimizing the area under the ROC curve (AUC), which is a more robust metric for imbalanced medical datasets.

However, AUC optimization alone can induce class dominance. To prevent this, we introduce a Cross-Entropy Difference Regularizer $\mathcal{L}_{Bal}$, which ensures that both classes in binary classification remain equally difficult to classify,

$$\mathcal{L}_{Bal} = \Big| \mathbb{E}_{(i|y_i=1)}[\mathcal{L}_{CE}(L_i, y_i)] - \mathbb{E}_{(i|y_i=0)}[\mathcal{L}_{CE}(L_i, y_i)] \Big| \quad (12)$$

where y_i corresponds to the ground-truth label for the i^{th} subject. Though cross-entropy is not used as a primary loss, it is used within this regularizer. This term acts as a stabilizer, penalizing imbalance across classes and thereby ensuring consistent performance even under severe label skew, generally found in medical datasets.

Composite Loss and Unified Optimization. The complete training process of the second stage classifier unites the above objectives into a single differentiable composite loss $\mathcal{L}_{total}$,

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{\text{Sinkhorn}} + \beta \mathcal{L}_{AUC} + \gamma \mathcal{L}_{Bal} \quad (13)$$

where α, β, and γ are the hyperparameters balancing the respective contributions of semantic alignment, discriminative ranking, and balance regularization. This integration ensures that the adapter learns to project the brain graph embeddings into the token space of a frozen LLM, the LLM learns to reason over structured brain information, and the classifier learns to make robust predictions.

4 Experimental Results

In this section, we conduct extensive experiments to evaluate the effectiveness of our proposed G-LaD framework in leveraging the latent reasoning capability of a frozen LLM for few-shot brain disorder classification. Our central research question is:

Can aligning graph-structured functional neuroimaging representations with a semantic space of a pre-trained langauge model enhance diagnostic accuracy and robustness under data-scarce conditions?

4.1 Dataset and Preprocessing

Experiments are performed on the Autism Brain Imaging Data Exchange (ABIDE) dataset (Di Martino et al. 2014), a benchmark multi-site collection of resting-state fMRI scans along with demographic and clinical metadata. ABIDE aggregates data from multiple international imaging centers, enabling the study of autism-related neurobiological patterns under diverse acquisition conditions and population variability. The task is formulated as a binary classification problem: distinguish individuals diagnosed with *Autism Spectrum Disorder (ASD)* from *Healthy Controls (HC)*.

Pre-processing. To construct the brain graph, we first preprocess the fMRI data using the DPARSF (Yan and Zang 2010) pipeline for standard resting-state fMRI correction and normalization. We then segment the brain into predefined ROIs based on the AAL atlas (Tzourio-Mazoyer et al. 2002), and extract the average time-series signal from each ROI to serve as node-level features.

4.2 Experimental Setup

Dataset Partitioning. In Stage 1, 80% of the dataset is allocated for pre-training. For Stage 2, we employ stratified partitioning, utilizing an 80:10:10 distribution ratio for the purposes of training, validation, and testing to ensure proportional representation of both classes across sets, as shown in Table 2. To simulate realistic data-scarce scenarios, the training is further subsampled to create 5% and 10% few-shot subsets, while preserving class balance.

Model Configuration. The Stage-1 encoder of G-LaD utilizes a 3-Layer GIN. Each layer applies linear transformation, followed by Batch Normalization and Dropout. This encoder architecture choice is based on the fact that GINs are amongst the most expressive MPNNs for capturing fine-grained topological information in brain connectivity graphs.

In Stage-2, we employed a frozen Llama-3-8B model (Grattafiori et al. 2024) within G-LaD. As discussed earlier, only the Stage-1, adapter (ϕ) and the classification head (ψ) are trainable, while the LLM backbone remains frozen throughout all experiments.

Training Details. The G-LaD framework is trained end-to-end using the steps defined in the earlier section. The losses are optimized using the AdamW optimizer with differentiable learning rates under a cosine learning rate schedule with a 10% warm-up phase. Stage-2 is trained for 30 epochs with a batch size of 16, and early stopping based on validation

Few-shot Setting	Method Type	Accuracy ↑	Specificity ↑	Sensitivity ↑	AUC ↑	F1 ↑
5%	GCN	$51.22_{\pm 2.78}$	$23.16_{\pm 13.87}$	**84.09**$_{\pm 8.62}$	$51.42_{\pm 6.41}$	$66.45_{\pm 2.32}$
	GAT	$50.95_{\pm 2.65}$	$25.10_{\pm 14.20}$	$82.30_{\pm 9.10}$	$51.10_{\pm 5.90}$	$65.80_{\pm 3.10}$
	GIN	$51.70_{\pm 2.40}$	$24.05_{\pm 13.00}$	$83.10_{\pm 8.90}$	$51.90_{\pm 6.00}$	$66.10_{\pm 2.80}$
	BrainGNN	$52.56_{\pm 3.10}$	$44.65_{\pm 27.56}$	$57.45_{\pm 22.67}$	$50.45_{\pm 4.67}$	$51.67_{\pm 6.87}$
	BrainNetCNN	$53.40_{\pm 3.20}$	$46.10_{\pm 25.40}$	$60.20_{\pm 21.30}$	$52.10_{\pm 4.90}$	$53.90_{\pm 7.50}$
	GCDA	$57.20_{\pm 3.05}$	$48.95_{\pm 23.70}$	$62.75_{\pm 19.85}$	$54.60_{\pm 4.50}$	$56.40_{\pm 7.10}$
	ALTER	$56.10_{\pm 2.95}$	$49.85_{\pm 22.90}$	$63.40_{\pm 19.20}$	$55.50_{\pm 4.30}$	$57.30_{\pm 6.80}$
	G-LaD (Ours)	**60.00**$_{\pm 2.18}$	**54.74**$_{\pm 19.29}$	$64.55_{\pm 20.05}$	**59.96**$_{\pm 2.16}$	**62.01**$_{\pm 9.71}$
10%	GCN	$53.78_{\pm 2.56}$	$43.63_{\pm 20.97}$	$57.67_{\pm 19.45}$	$58.00_{\pm 4.16}$	$55.45_{\pm 6.54}$
	GAT	$54.10_{\pm 2.70}$	$44.20_{\pm 18.90}$	$58.30_{\pm 18.10}$	$58.40_{\pm 4.30}$	$55.90_{\pm 6.20}$
	GIN	$54.00_{\pm 2.35}$	$42.95_{\pm 19.10}$	$59.10_{\pm 18.90}$	$58.10_{\pm 4.45}$	$56.10_{\pm 6.40}$
	BrainGNN	$54.45_{\pm 4.84}$	$33.45_{\pm 17.56}$	$69.76_{\pm 24.56}$	$57.43_{\pm 5.10}$	$57.45_{\pm 15.45}$
	BrainNetCNN	$55.80_{\pm 3.90}$	$36.80_{\pm 18.10}$	$71.95_{\pm 22.10}$	$58.95_{\pm 5.20}$	$58.90_{\pm 8.90}$
	GCDA	$59.80_{\pm 3.50}$	$45.50_{\pm 17.85}$	$71.20_{\pm 20.50}$	$60.70_{\pm 4.80}$	$61.20_{\pm 8.40}$
	ALTER	$58.60_{\pm 3.20}$	$41.10_{\pm 18.00}$	$70.40_{\pm 19.90}$	$61.80_{\pm 5.00}$	$62.40_{\pm 8.10}$
	G-LaD (Ours)	**64.63**$_{\pm 3.11}$	**54.73**$_{\pm 19.90}$	**73.18**$_{\pm 17.34}$	**63.57**$_{\pm 6.12}$	**68.24**$_{\pm 6.19}$

Table 3: **Few-shot Autism Classification Performance on ABIDE.** Classification performance under 5% and 10% few-shot settings across traditional GNN architectures (GCN, GAT, GIN), brain disorder–specific baselines (BrainGNN, BrainNetCNN, GCDA, ALTER), and our proposed G-LaD framework. G-LaD consistently achieves superior and more balanced performance across all metrics in both few-shot regimes, highlighting its robustness and generalization in scarce-data scenarios.

Few-shot Setting	Ablation Variant	Accuracy ↑	Specificity ↑	Sensitivity ↑	AUC ↑	F1 ↑
5%	Only Stage-1	$54.88_{\pm 2.40}$	**92.11**$_{\pm 29.62}$	$22.73_{\pm 28.92}$	$50.31_{\pm 1.52}$	$35.09_{\pm 15.65}$
	w/o $\mathcal{L}_{Sinkhorn}$ & $\mathcal{L}_{Bal}$	$52.68_{\pm 4.10}$	$41.31_{\pm 26.64}$	$62.50_{\pm 19.62}$	$49.73_{\pm 6.31}$	$57.38_{\pm 7.61}$
	w/o $\mathcal{L}_{Sinkhorn}$	$54.07_{\pm 5.28}$	$47.66_{\pm 16.03}$	$59.60_{\pm 19.01}$	$53.85_{\pm 4.60}$	$56.93_{\pm 10.41}$
	w/o $\mathcal{L}_{Bal}$	$56.25_{\pm 4.10}$	$41.12_{\pm 12.81}$	**69.32**$_{\pm 12.57}$	$56.02_{\pm 4.27}$	$59.86_{\pm 6.10}$
	G-LaD (Ours)	**60.00**$_{\pm 2.18}$	$54.74_{\pm 19.29}$	$64.55_{\pm 20.05}$	**59.96**$_{\pm 2.16}$	**62.01**$_{\pm 9.71}$
10%	Only Stage-1	$56.10_{\pm 2.61}$	**68.42**$_{\pm 19.37}$	$45.45_{\pm 16.56}$	$52.63_{\pm 5.11}$	$53.17_{\pm 7.54}$
	w/o $\mathcal{L}_{Sinkhorn}$ & $\mathcal{L}_{Bal}$	$53.17_{\pm 4.01}$	$56.84_{\pm 21.02}$	$50.00_{\pm 20.89}$	$51.41_{\pm 6.19}$	$51.54_{\pm 11.73}$
	w/o $\mathcal{L}_{Sinkhorn}$	$55.61_{\pm 1.39}$	$38.95_{\pm 19.29}$	$70.00_{\pm 14.41}$	$51.78_{\pm 5.95}$	$62.39_{\pm 4.26}$
	w/o $\mathcal{L}_{Bal}$	$57.81_{\pm 2.22}$	$41.05_{\pm 13.50}$	$72.27_{\pm 12.59}$	$57.17_{\pm 2.68}$	$64.37_{\pm 5.11}$
	G-LaD (Ours)	**64.63**$_{\pm 3.11}$	$54.73_{\pm 19.90}$	**73.18**$_{\pm 17.34}$	**63.57**$_{\pm 6.12}$	**68.24**$_{\pm 6.19}$

Table 4: **Ablation study under 5% and 10% few-shot settings.** Each configuration isolates the contribution of core components in the G-LaD framework. Highlighted (yellow) cells indicate the best performance within each metric. The complete **G-LaD** model consistently achieves superior performance across most metrics, demonstrating the complementary importance of $\mathcal{L}_{Sinkhorn}$ and $\mathcal{L}_{Bal}$ for balanced, multimodal alignment.

Balanced Accuracy with a patience of 5. All training is performed on one Nvidia A100 GPU. Also, each experiment is conducted over 5 random seeds with mean and standard deviations reported in Tables 3 and 4.

Metrics. Given the clinical relevance of neurological disorder classification, the performance of G-LaD is evaluated using a combination of conventional Machine Learning metrics and medically interpretable indicators. Specifically, we report Accuracy (ACC), F-1 Score, and the Area Under the Receiver Operating Characteristics Curve(AUC), along with two clinically significant metrics: Sensitivity and Specificity, which respectively measure the model's ability to correctly classify diseased and healthy subjects.

4.3 Baselines

To evaluate the effectiveness of our proposed model, we compared its performance with several established graph-based neural architectures. We consider the Graph Convolutional Network (GCN) as a fundamental baseline for graph representation learning and further compare our model with its variants, Graph Attention Network (GAT) and Graph Isomorphism Network (GIN), to assess the impact of different graph convolution operations.

BrainGNN (Li et al. 2021) introduces an interpretable framework for fMRI analysis, incorporating ROI-aware graph convolution and ROI-selection pooling layers to identify salient brain regions and communities associated with neurological biomarkers. In contrast, BrainNetCNN (Kawahara et al. 2017) is a CNN designed specifically for structural brain connectomes derived from diffusion MRI. It employs three novel filter types, edge-to-edge, edge-to-node, and node-to-graph, to capture the hierarchical organization of brain connectivity patterns. GCDA (Wang et al. 2025) uses diffusion-based contrastive augmentation for self-supervised representation learning. ALTER (Yu et al. 2024) introduces a long-

range attention-based graph learning framework designed for brain connectome analysis.

These methods together form a comprehensive set of baselines to benchmark the representational and predictive capabilities of our model in brain network analysis.

4.4 Results

Table 3 summarizes the comparative performance between the G-LaD framework and baseline methods. Across both the 5% and 10% label regimes, G-LaD consistently and significantly outperforms all competing approaches. At the 5% label rate, G-LaD surpasses the strongest baseline, GCDA, by approximately 2.8% in accuracy and exhibits balanced sensitivity–specificity trade-offs, indicating robust learning of discriminative and generalizable representations even in low-data conditions. Similarly, under the 10% label regime, G-LaD demonstrates a substantial performance gain of nearly 5% in accuracy and 7% in F1-score compared to the GCDA.

4.5 Ablation Study

In order to assess the contribution of individual components of our proposed G-LaD framework, we performed a comprehensive ablation study under both 5% and 10% few-shot settings. Specifically, we evaluate four different ablations of our proposed framework:

- Using only the Stage-1 graph encoder without multimodal alignment to assess the independent contribution of structural graph representations.
- Excluding $\mathcal{L}_{Sinkhorn}$ to evaluate the role of cross-modal distribution matching between graph embeddings and LLM's token space.
- Excluding $\mathcal{L}_{Bal}$ to examine the influence of the class-balance constraint on mitigating label imbalance during training.
- Finally, removing both $\mathcal{L}_{Sinkhorn}$ and $\mathcal{L}_{Bal}$ to determine the combined effect of omitting semantic alignment and balance regularization.

As is evident in Table 4, the best set of balanced results for both 5% and 10% settings were achieved only when all the components of G-LaD were utilized for classification.

5 Conclusion

This work provides a first attempt to integrate knowledge from a pretrained LLMs with a GNN-based brain graph encoder for functional connectivity analysis. We introduce G-LaD, a two-stage framework that aligns multimodal structural and semantic representations to improve autism identification. Our results demonstrate the effectiveness of the proposed approach, particularly in label-scarce few-shot settings. In future work, we plan to extend this framework to additional neurological disorders and develop LLM-driven interpretability tools to further support clinical decision-making and assist healthcare professionals.

Acknowledgments

This work was supported by the Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, through the provision of GPU resources and research infrastructure.

References

Abraham, A.; Milham, M. P.; Di Martino, A.; Craddock, R. C.; Samaras, D.; Thirion, B.; and Varoquaux, G. 2017. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *NeuroImage*, 147: 736–745.

Bessadok, A.; Mahjoub, M. A.; and Rekik, I. 2022. Graph neural networks in network neuroscience. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5): 5833–5848.

Brannon, W.; Kang, W.; Fulay, S.; Jiang, H.; Roy, B.; Roy, D.; and Kabbara, J. 2023. Congrat: Self-supervised contrastive pretraining for joint graph and text embeddings. *arXiv preprint arXiv:2305.14321*.

Chen, R.; Zhao, T.; Jaiswal, A.; Shah, N.; and Wang, Z. 2024a. Llaga: Large language and graph assistant. *arXiv preprint arXiv:2402.08170*.

Chen, Z.; Mao, H.; Li, H.; Jin, W.; Wen, H.; Wei, X.; Wang, S.; Yin, D.; Fan, W.; Liu, H.; et al. 2024b. Exploring the potential of large language models (llms) in learning on graphs. *ACM SIGKDD Explorations Newsletter*, 25(2): 42–61.

Di Martino, A.; Yan, C.-G.; Li, Q.; Denio, E.; Castellanos, F. X.; Alaerts, K.; Anderson, J. S.; Assaf, M.; Bookheimer, S. Y.; Dapretto, M.; et al. 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6): 659–667.

Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Vaughan, A.; et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Huang, J.; Zhang, X.; Mei, Q.; and Ma, J. 2023. Can llms effectively leverage graph structural information through prompts, and why? *arXiv preprint arXiv:2309.16595*.

Jin, B.; Liu, G.; Han, C.; Jiang, M.; Ji, H.; and Han, J. 2024. Large language models on graphs: A comprehensive survey. *IEEE Transactions on Knowledge and Data Engineering*.

Kan, X.; Cui, H.; Lukemire, J.; Guo, Y.; and Yang, C. 2022a. Fbnetgen: Task-aware gnn-based fmri analysis via functional brain network generation. In *International conference on medical imaging with deep learning*, 618–637. PMLR.

Kan, X.; Dai, W.; Cui, H.; Zhang, Z.; Guo, Y.; and Yang, C. 2022b. Brain network transformer. *Advances in Neural Information Processing Systems*, 35: 25586–25599.

Kawahara, J.; Brown, C. J.; Miller, S. P.; Booth, B. G.; Chau, V.; Grunau, R. E.; Zwicker, J. G.; and Hamarneh, G. 2017. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146: 1038–1049.

Lee, S. H.; Yu, D.; Bachman, A. H.; Lim, J.; and Ardekani, B. A. 2014. Application of fused lasso logistic regression to the study of corpus callosum thickness in early Alzheimer's disease. *Journal of neuroscience methods*, 221: 78–84.

Li, X.; Zhou, Y.; Dvornek, N.; Zhang, M.; Gao, S.; Zhuang, J.; Scheinost, D.; Staib, L. H.; Ventola, P.; and Duncan, J. S. 2021. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical Image Analysis*, 74: 102233.

Li, Y.; Ding, K.; and Lee, K. 2023. Grenade: Graph-centric language model for self-supervised representation learning on text-attributed graphs. *arXiv preprint arXiv:2310.15109*.

Liu, S.; Nie, W.; Wang, C.; Lu, J.; Qiao, Z.; Liu, L.; Tang, J.; Xiao, C.; and Anandkumar, A. 2023. Multi-modal molecule structure–text model for text-based retrieval and editing. *Nature Machine Intelligence*, 5(12): 1447–1457.

Logothetis, N. K.; and Wandell, B. A. 2004. Interpreting the BOLD signal. *Annu. Rev. Physiol.*, 66: 735–769.

Mitchell, T. M.; Hutchinson, R.; Niculescu, R. S.; Pereira, F.; Wang, X.; Just, M.; and Newman, S. 2004. Learning to decode cognitive states from brain images. *Machine learning*, 57(1): 145–175.

Morra, J. H.; Tu, Z.; Apostolova, L. G.; Green, A. E.; Toga, A. W.; and Thompson, P. M. 2009. Comparison of AdaBoost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation. *IEEE transactions on medical imaging*, 29(1): 30–43.

Oneto, L.; Donini, M.; Luise, G.; Ciliberto, C.; Maurer, A.; and Pontil, M. 2020. Exploiting mmd and sinkhorn divergences for fair and transferable representation learning. *Advances in Neural Information Processing Systems*, 33: 15360–15370.

Peng, L.; Wang, N.; Xu, J.; Zhu, X.; and Li, X. 2022. GATE: Graph CCA for temporal self-supervised learning for label-efficient fMRI analysis. *IEEE Transactions on Medical Imaging*, 42(2): 391–402.

Su, B.; Du, D.; Yang, Z.; Zhou, Y.; Li, J.; Rao, A.; Sun, H.; Lu, Z.; and Wen, J.-R. 2022. A molecular multimodal foundation model associating molecule graphs with natural language. *arXiv preprint arXiv:2209.05481*.

Tang, J.; Yang, Y.; Wei, W.; Shi, L.; Su, L.; Cheng, S.; Yin, D.; and Huang, C. 2024. Graphgpt: Graph instruction tuning for large language models. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 491–500.

Tang, X.; Qin, Y.; Wu, J.; Zhang, M.; Zhu, W.; and Miller, M. I. 2016. Shape and diffusion tensor imaging based integrative analysis of the hippocampus and the amygdala in Alzheimer's disease. *Magnetic resonance imaging*, 34(8): 1087–1099.

Tzourio-Mazoyer, N.; Landeau, B.; Papathanassiou, D.; Crivello, F.; Etard, O.; Delcroix, N.; Mazoyer, B.; and Joliot, M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1): 273–289.

Vincent, P.; Larochelle, H.; Bengio, Y.; and Manzagol, P.-A. 2008. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, 1096–1103.

Wang, P.; Zhang, H.; He, Z.; Peng, Z.; and Yuan, Y. 2024. ftspl: Enhancing brain analysis with fmri-text synergistic prompt learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 564–574. Springer.

Wang, X.; Fang, Y.; Wang, Q.; Yap, P.-T.; Zhu, H.; and Liu, M. 2025. Self-supervised graph contrastive learning with diffusion augmentation for functional MRI analysis and brain disorder detection. *Medical image analysis*, 101: 103403.

Wee, C.-Y.; Yap, P.-T.; Zhang, D.; Denny, K.; Browndyke, J. N.; Potter, G. G.; Welsh-Bohmer, K. A.; Wang, L.; and Shen, D. 2012. Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage*, 59(3): 2045–2056.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.

Wen, D.; Wei, Z.; Zhou, Y.; Li, G.; Zhang, X.; and Han, W. 2018. Deep learning methods to process fmri data and their application in the diagnosis of cognitive impairment: a brief overview and our opinion. *Frontiers in neuroinformatics*, 12: 23.

Wen, Z.; and Fang, Y. 2023. Augmenting low-resource text classification with graph-grounded pre-training and prompting. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 506–516.

Wen, Z.; and Fang, Y. 2024. Prompt tuning on graph-augmented low-resource text classification. *IEEE Transactions on Knowledge and Data Engineering*.

Xu, J.; He, K.; Tang, Y.; Li, W.; Lan, M.; Dong, X.; Ke, Y.; and Feng, M. 2025. BrainPrompt: Multi-level Brain Prompt Enhancement for Neurological Condition Identification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 172–182. Springer.

Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*.

Yan, C.; and Zang, Y. 2010. DPARSF: a MATLAB toolbox for" pipeline" data analysis of resting-state fMRI. *Frontiers in systems neuroscience*, 4: 1377.

Yu, S.; Jin, S.; Li, M.; Sarwar, T.; and Xia, F. 2024. Long-range brain graph transformer. *Advances in Neural Information Processing Systems*, 37: 24472–24495.

Zhao, J.; Qu, M.; Li, C.; Yan, H.; Liu, Q.; Li, R.; Xie, X.; and Tang, J. 2022. Learning on large-scale text-attributed graphs via variational inference. *arXiv preprint arXiv:2210.14709*.

Zou, T.; Yu, L.; Huang, Y.; Sun, L.; and Du, B. 2023. Pre-training language models with text-attributed heterogeneous graphs. *arXiv preprint arXiv:2310.12580*.