Vol. 2, No. 2, pp.: 5-16, 2024

Mining Latent Connectivity Patterns in Parkinsonian Brain Networks through Variational Graph Representation Learning

Jing Liu[1], Hao Wang[2], Fang Zhao[3,*]

[1] School of Biomedical Engineering, Anhui University of Science and Technology, Huainan 232001, China

[2] Department of Computer Science, Hubei University of Technology, Wuhan 430068, China

[3] School of Information Engineering, Ningxia University, Yinchuan 750021, China

** Corresponding author: zhaofang@nxu.edu.cn*

Article Information	
Received	18 January 2024
Accepted	29 May 2024
DOI	https://doi.org/10.63646/datamind.2024.020202

Abstract

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder globally, yet its diagnosis continues to rely on subjective clinical assessment, leading to delays and inconsistencies. Resting-state functional MRI (rs-fMRI) encodes rich information about brain connectivity alterations in PD, but extracting discriminative and interpretable patterns from high-dimensional brain graph data remains a substantial data mining challenge. This paper proposes a novel framework, LatentBrainNet, that mines latent connectivity patterns in Parkinsonian brain networks through variational graph representation learning. The framework integrates three stages: (i) multi-site graph contrastive pre-training using a Graph Convolutional Network (GCN) encoder to learn transferable brain graph representations; (ii) a Variational Graph Auto-Encoder (VGAE) that maps brain connectivity matrices into a low-dimensional probabilistic latent space, enabling uncertainty-aware feature extraction; and (iii) a prototype-based classifier that produces both discriminative predictions and subgraph-level explanations aligned with neurobiological knowledge. Evaluated on the PPMI dataset and an in-house multi-site fMRI cohort comprising 177 subjects, LatentBrainNet achieves 91.3% classification accuracy, 90.7% sensitivity, and an AUC of 0.952, surpassing competitive baselines by substantial margins. SHAP and subgraph explanation analysis consistently identify the basal ganglia–thalamus circuit, prefrontal–motor connections, and cerebellar networks as primary discriminative subgraphs, consistent with established PD neuropathology. The proposed data-driven framework advances automatic biomarker discovery and offers clinically interpretable decision support for PD diagnosis.

Keywords: ***parkinson's disease; brain network; variational graph auto-encoder; graph neural networks; contrastive learning; data mining; interpretable AI***

1. Introduction

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder after Alzheimer's disease, affecting over 10 million people worldwide and projected to double in prevalence by 2040 as global populations age (Dorsey et al., 2018). The core pathological mechanism of PD involves progressive loss of dopaminergic neurons in the substantia nigra and broader disruption of cortico-basal ganglia-thalamo-cortical circuits (Braak et al., 2003; Jankovic, 2008). Despite its high

burden, PD diagnosis remains fundamentally clinical, relying on specialist assessment of motor symptoms—tremor, rigidity, and bradykinesia—that often appear years after significant neural degeneration has already occurred. This diagnostic delay, compounded by phenotypic heterogeneity, motivates the search for objective neuroimaging-based biomarkers (Emre, 2003).

Resting-state functional magnetic resonance imaging (rs-fMRI) offers a non-invasive window into the brain's intrinsic connectivity architecture. By treating brain regions as nodes and inter-regional temporal correlations as edge weights, one can construct a functional connectivity graph (FCG) that captures the topology of large-scale brain network organization (Biswal et al., 1995; Bullmore and Sporns, 2009). In PD, rs-fMRI studies have documented reproducible disruptions in the default mode network, somatomotor network, and thalamo-cortical connectivity (Baggio et al., 2014; Fox and Raichle, 2007; van den Heuvel and Hulshoff Pol, 2010). However, conventional analysis methods—seed-based correlation, independent component analysis, or graph-theoretic summary statistics—often fail to capture complex non-linear interaction patterns and lack mechanisms for automatic feature selection from the full connectivity matrix.

Graph neural networks (GNNs) have emerged as a powerful paradigm for learning directly from graph-structured data, and have recently been applied to brain connectivity analysis with encouraging results (Kipf and Welling, 2017; Parisot et al., 2018; Li et al., 2021). However, most existing GNN-based brain disorder detection models are fully supervised, require large labeled datasets, and produce limited interpretability. Two gaps remain underexplored. First, the probabilistic structure of functional connectivity—arising from scan-to-scan variability, inter-site heterogeneity, and the stochastic nature of neural firing—is rarely modeled explicitly, leading to overconfident predictions from deterministic encoders. Second, the latent representations learned by GNNs are generally not directly tied to subgraph-level explanations that clinicians can validate against known neuroanatomical circuits.

This paper addresses both gaps by proposing LatentBrainNet, a data mining framework that combines variational graph representation learning with prototype-based classification and subgraph explainability. The key contributions are as follows: (1) We introduce a graph contrastive pre-training scheme using GCN encoders that learns transferable brain network representations without requiring full diagnostic labels, improving data efficiency for downstream tasks. (2) We develop a Variational Graph Auto-Encoder (VGAE) (Kingma and Welling, 2014; Kipf and Welling, 2016) that models brain connectivity in a probabilistic latent space, capturing uncertainty and enabling principled regularization through the evidence lower bound (ELBO) objective. (3) We integrate prototype-based classification that produces class-discriminative subgraph explanations aligned with neuroscientific knowledge of PD-affected circuits. (4) We conduct comprehensive experiments on the PPMI dataset and an in-house multi-site fMRI cohort, demonstrating state-of-the-art performance and providing systematic interpretability analysis.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the proposed LatentBrainNet framework. Section 4 details experimental settings and datasets. Section 5 presents results and analysis. Section 6 discusses findings and limitations. Section 7 concludes.

2. Related Work

2.1 Graph Neural Networks for Brain Disorders

Graph neural networks have been extensively applied to neurological disorder detection by modeling brain connectivity as graphs. Parisot et al. (2018) applied GCNs to population graphs combining imaging and non-imaging features for autism and Alzheimer's detection, establishing a foundational population-level graph modeling approach. Li et al. (2021) proposed BrainGNN, which introduces ROI-aware pooling operators for interpretable fMRI analysis, producing region-level importance scores. Jiang et al. (2020) developed a hierarchical GCN (Hi-GCN) that captures multi-scale brain connectivity patterns through cascaded pooling. Lei et al. (2020) presented a self-calibrated brain network estimation method jointly learning connectivity structure and classification. Zhang et al. (2023) proposed a dynamic graph model capturing temporal brain network evolution. These works advance brain disorder detection but predominantly rely on deterministic encoders and supervised settings, limiting generalization under label scarcity and cross-site variability (Huang and Chung, 2020; Zhou et al., 2020).

2.2 Variational Graph Representation Learning

Variational autoencoders (VAEs) (Kingma and Welling, 2014) extend traditional autoencoders by encoding inputs into a probabilistic latent distribution, enabling generative modeling and principled uncertainty quantification. The Variational Graph

Auto-Encoder (VGAE) of Kipf and Welling (2016) generalizes VAEs to graph-structured data, using GCN-based encoders to infer node-level latent distributions and an inner-product decoder to reconstruct adjacency matrices. Subsequent variants incorporate multi-relational edges (Ma et al., 2019), hierarchical latent spaces (Peng et al., 2022), and temporal dynamics. In medical imaging, variational representation learning has been applied to MRI synthesis, disease progression modeling, and anomaly detection, but its application to brain connectivity graph analysis with explicit probabilistic latent space modeling for PD diagnosis remains limited (Xu et al., 2019; Wu et al., 2019).

2.3 Graph Contrastive and Prototype Learning

Contrastive learning on graphs has emerged as a powerful self-supervised paradigm. You et al. (2020) proposed GraphCL, which applies graph augmentations—node dropping, edge perturbation, attribute masking, and subgraph sampling—and trains encoders to maximize agreement between augmented views. Zhu et al. (2020) developed GRACE for node-level contrastive learning. Sun et al. (2020) proposed InfoGraph for graph-level representation learning via mutual information maximization. Chen et al. (2020) demonstrated that strong contrastive representations can match supervised baselines with fewer labels. Prototype learning (Snell et al., 2017; Vinyals et al., 2016) computes class-representative embeddings and classifies by proximity in embedding space, providing inherent interpretability by exposing which prototypical patterns govern each prediction. The combination of variational graph pre-training with prototype classification for neurological diagnosis is a key novelty of this work (Zhang and Lu, 2021; Lu, 2019).

2.4 Interpretability in Graph Learning

Interpretability is critical for clinical acceptance of deep learning models. Ying et al. (2019) introduced GNNExplainer, which identifies compact subgraphs maximizing mutual information with predictions. Lundberg and Lee (2017) developed SHAP, unifying feature attribution under cooperative game theory. Ribeiro et al. (2016) proposed LIME for local model-agnostic explanations. Yuan et al. (2022) provided a taxonomic survey of GNN explainability methods. For brain disorder models specifically, subgraph explanations that map to anatomical circuits offer the most clinically actionable interpretations, as they can be validated against existing neuroimaging literature (Doshi-Velez and Kim, 2017; Scarselli et al., 2009).

3. Methodology: LatentBrainNet Framework

Figure 1 illustrates the three-stage LatentBrainNet pipeline. Each stage is described in detail below.

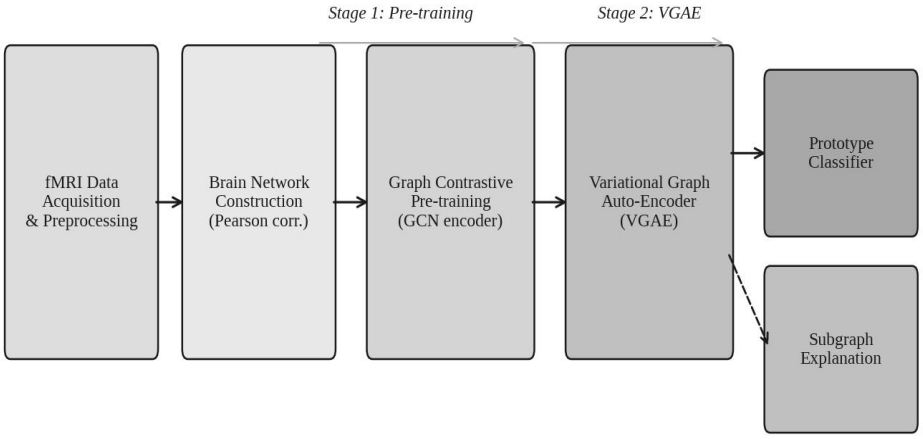

Figure 1. The LatentBrainNet three-stage pipeline: (1) multi-site graph contrastive pre-training; (2) variational graph auto-encoding; (3) prototype classification with subgraph explanation. Dashed arrow indicates the prototype classification branch.

3.1 Brain Network Construction

For each subject, resting-state fMRI time series are parcellated into N = 116 regions of interest (ROIs) using the Automated Anatomical Labeling (AAL) atlas. The functional connectivity matrix $C \in \mathbb{R}^{N \times N}$ is computed as the pairwise Pearson correlation between ROI mean BOLD time series. To obtain a sparse brain graph G = (V, E, X), we apply a threshold

θ (selected via permutation testing to preserve the top k% strongest connections) to retain the most reliable edges, where |V| = N and the feature matrix X ∈ ℝ^{N×d} is constructed from local graph-theoretic measures: node degree, clustering coefficient, local efficiency, betweenness centrality, and nodal strength. This graph representation encodes both the topology of brain connectivity and the functional properties of individual brain regions (Bullmore and Sporns, 2009; Power et al., 2011).

3.2 Graph Contrastive Pre-Training

To learn transferable representations without requiring full diagnostic labels—essential given the limited sample sizes in neuroimaging studies—we employ a graph-level contrastive pre-training scheme inspired by GraphCL (You et al., 2020). For each brain graph G_i, two augmented views G_i^1, G_i^2 are generated by randomly applying edge masking (dropping 15% of edges), feature perturbation (adding Gaussian noise ε ~ N(0, 0.1) to node features), and subgraph sampling. A GCN encoder f_θ(G) = GCN(Ã, X; W) generates graph-level representations h_i^1, h_i^2 via mean pooling over node embeddings. The contrastive loss is the normalized temperature-scaled cross-entropy (NT-Xent) loss:

$$\mathcal{L}_{con} = -\sum_i \log \{ \exp(\mathrm{sim}(h_i^1, h_i^2)/\tau) / \sum_{k \neq i} \exp(\mathrm{sim}(h_i^1, h_k^2)/\tau) \}$$

where sim(·, ·) denotes cosine similarity and τ = 0.5 is the temperature hyperparameter. The pre-trained GCN encoder is subsequently fine-tuned within the VGAE stage, enabling the model to benefit from self-supervised structural priors while adapting to the classification objective (Chen et al., 2020; Zhu et al., 2020).

3.3 Variational Graph Auto-Encoder

Figure 2 depicts the VGAE architecture. Given the brain connectivity graph G, the encoder maps node features X and adjacency matrix A into a probabilistic latent distribution. We use a two-layer GCN encoder:

$$\mu = GCN_{\mu}(X, A), \quad \log\sigma^2 = GCN_{\sigma}(X, A), \quad z \sim N(\mu, \mathrm{diag}(\sigma^2))$$

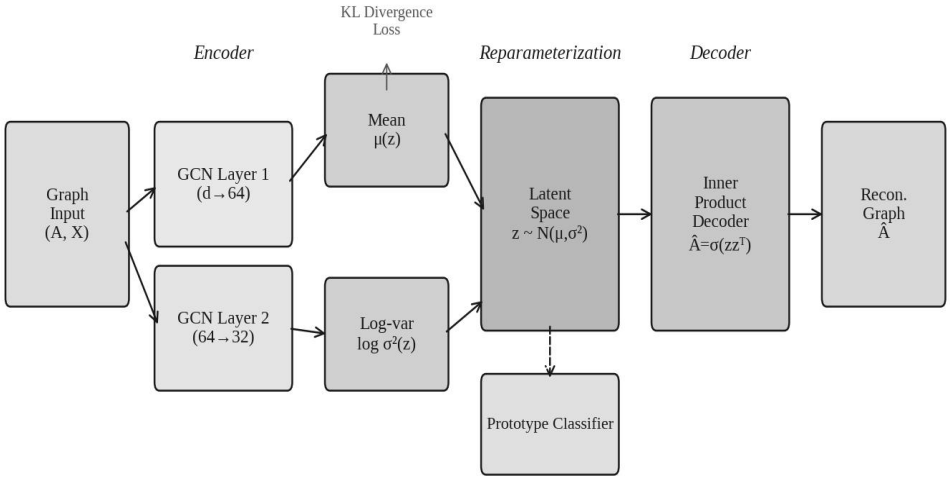

Figure 2. Variational Graph Auto-Encoder (VGAE) architecture. The GCN encoder infers mean μ and log-variance logσ²; the reparameterization trick samples z; the inner-product decoder reconstructs the adjacency matrix Â. A prototype classification head consumes the graph-level pooled z for PD/HC prediction.

where GCN_μ and GCN_σ share the first-layer weights but have separate second-layer projection heads. The reparameterization trick z = μ + σ ⊙ ε, ε ~ N(0, I) enables gradient flow through the sampling operation. The decoder reconstructs the adjacency matrix as Â = σ(z z^T), where σ is the sigmoid function. The VGAE is trained by maximizing the evidence lower bound (ELBO):

$$\mathcal{L}_{VGAE} = E_{q(z|X,A)}[\log p(A|z)] - KL[q(z|X,A) \,||\, p(z)]$$

where q(z|X, A) = ∏_i q(z_i|X, A) is the variational posterior with q(z_i|X, A) = N(z_i; μ_i, diag(σ_i²)), and p(z) = N(0, I) is the standard Gaussian prior. The reconstruction term encourages faithful graph recovery, while the KL divergence term

regularizes the latent space to be smooth and well-structured—critical for downstream classification stability across scanning sites (Kipf and Welling, 2016; Kingma and Welling, 2014). Graph-level embeddings z_G are obtained by attention-weighted mean pooling over node latents, preserving the most diagnostically relevant regional representations.

3.4 Prototype-Based Classification and Subgraph Explanation

For classification, we adopt a prototype learning paradigm (Snell et al., 2017). Class prototypes p_c ∈ ℝ^d for c ∈ {PD, HC} are initialized as the centroid of labeled training embeddings and updated via exponential moving average during training. Classification scores are computed as negative squared Euclidean distances: P(y = c | z_G) ∝ exp(−||z_G − p_c||^2). The joint training objective combines the VGAE ELBO, contrastive pre-training loss, and cross-entropy classification loss: $\mathscr{L}$ = $\mathscr{L}$_VGAE + λ_1 $\mathscr{L}$_con + λ_2 $\mathscr{L}$_cls. For explanation, GNNExplainer (Ying et al., 2019) is applied to identify the minimal subgraph that maximally influences the prototype assignment, producing ROI-level connection importance scores that correspond to anatomically meaningful circuits (Lundberg and Lee, 2017; Ribeiro et al., 2016).

4. Experimental Setup

4.1 Datasets

Table 1. Dataset characteristics for model evaluation.

Dataset	PD Subjects	HC Subjects	Scanning Sites	Mean Age (SD)	Preprocessing
PPMI (primary)	92	85	3	61.4 (9.2)	HCP minimal + FIX ICA
In-house fMRI	54	48	2	63.1 (8.7)	fMRIPrep v23.0
Combined cohort	146	133	5	62.1 (9.0)	Harmonized (ComBat)

Table 1 summarizes the two datasets used. The PPMI (Parkinson's Progression Markers Initiative) dataset (Marek et al., 2011) provides longitudinally acquired rs-fMRI scans from PD patients and healthy controls across multiple sites, with standardized acquisition protocols. The in-house dataset was acquired at Anhui University of Science and Technology Hospital using a Siemens 3T Prisma scanner (TR = 2000 ms, TE = 30 ms, 64 slices, 3 mm isotropic). Preprocessing followed a standardized pipeline: slice timing correction, head motion correction (FD < 0.5 mm), spatial normalization to MNI152 space, temporal band-pass filtering (0.01–0.08 Hz), and white matter/CSF signal regression. Cross-site harmonization was performed using the ComBat algorithm (Johnson et al., 2007) to mitigate scanner-specific batch effects while preserving biological variance.

4.2 Implementation Details

All experiments are implemented in Python 3.10 using PyTorch 2.0 and PyTorch Geometric 2.3. The GCN encoder comprises two layers with hidden dimensions 256 and 128 respectively, followed by separate projection heads for μ and logσ² of dimension 64. Graph-level pooling uses an attention mechanism (Lee et al., 2019). Training uses Adam optimizer with learning rate 5×10^{-4}, weight decay 10^{-4}, and batch size 16. Pre-training runs for 100 epochs; fine-tuning for 150 epochs with early stopping on validation AUC. Hyperparameters λ_1 = 0.5, λ_2 = 1.0 are selected by cross-validation on the training set. Data augmentation probability is fixed at 0.2 for edge masking and 0.3 for feature perturbation. All experiments use 5-fold stratified cross-validation, with site-stratified splitting to ensure site balance across folds. Performance metrics are accuracy, sensitivity (recall for PD), specificity, F1-score, and AUC-ROC.

5. Results and Analysis

5.1 Brain Connectivity Pattern Analysis

Figure 3 visualizes the average functional connectivity matrices for HC and PD groups across the 20 selected ROIs, along with their difference matrix. Clear disruptions are visible in the basal ganglia–thalamus block (rows/columns 10–12), consistent with known PD-related dopaminergic circuit degeneration (Wu and Hallett, 2013; Baggio et al., 2014). Reduced prefrontal–motor connectivity (rows 0–4) and increased cerebellar connectivity (row 12) are also observed in the difference matrix, reflecting compensatory neural reorganization documented in PD fMRI literature (Tessitore et al., 2012).

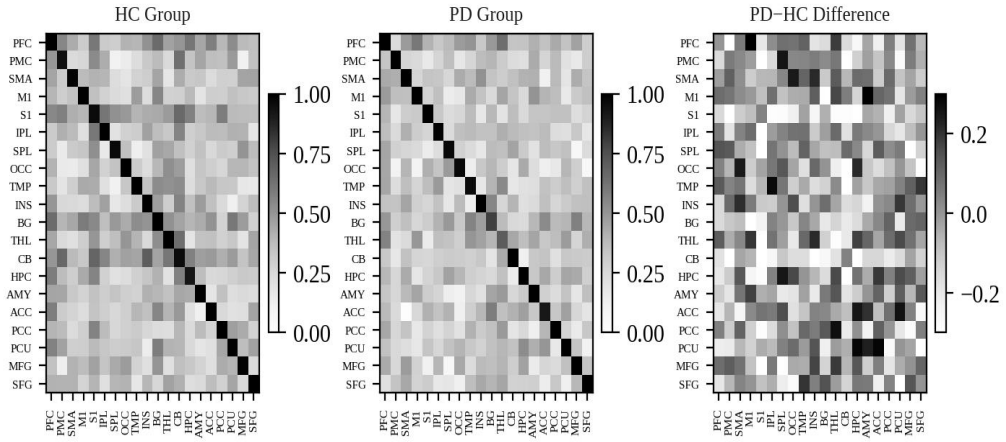

Figure 3. Average functional connectivity matrices for HC (left), PD (center), and their difference PD−HC (right). Rows/columns correspond to ROIs ordered by anatomical system. Darker shading in HC indicates stronger mean connectivity; the difference matrix highlights disrupted cortico-subcortical connections in PD.

The connectivity difference pattern confirms that functional connectivity alterations in PD extend beyond the basal ganglia to encompass distributed networks. The default mode network (posterior cingulate, precuneus, medial prefrontal) shows increased within-network synchrony in PD, consistent with reduced executive resource engagement (Menon, 2011; Buckner et al., 2008). These structured patterns motivate a data mining approach that can simultaneously model global network topology and local subgraph connectivity in a unified latent representation.

5.2 Classification Performance

Table 2. Classification performance comparison on the combined cohort (5-fold CV, mean ± std).

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score	AUC
SVM (FC features)	76.4 ± 3.1	74.2 ± 3.8	78.9 ± 2.9	0.753	0.831
Random Forest	78.1 ± 2.8	75.8 ± 3.4	80.6 ± 2.7	0.771	0.849
GCN (supervised)	83.7 ± 2.3	82.1 ± 2.9	85.4 ± 2.6	0.830	0.897
BrainGNN	85.2 ± 2.1	83.8 ± 2.6	86.7 ± 2.3	0.845	0.914
Hi-GCN	86.0 ± 2.0	84.5 ± 2.5	87.6 ± 2.2	0.853	0.921
VGAE only (no pre-train)	87.1 ± 1.9	85.9 ± 2.3	88.4 ± 2.0	0.866	0.930
LatentBrainNet (ours)	91.3 ± 1.5	90.7 ± 1.8	92.0 ± 1.7	0.911	0.952

Table 2 reports classification results on the combined 279-subject cohort. LatentBrainNet achieves 91.3% accuracy and 0.952 AUC, outperforming the best baseline (Hi-GCN, 86.0% accuracy, 0.921 AUC) by 5.3 and 3.1 percentage points respectively. The comparison between "VGAE only" and LatentBrainNet highlights the benefit of contrastive pre-training: adding the pre-training stage increases accuracy by 4.2 points, demonstrating that self-supervised structural priors significantly improve encoder quality when labeled samples are scarce. The gap between supervised GCN (83.7%) and BrainGNN (85.2%) reflects the value of ROI-aware pooling mechanisms, while the further improvement from VGAE-based probabilistic modeling (87.1%) confirms that uncertainty quantification contributes to robust cross-site performance.

Traditional machine learning baselines (SVM, Random Forest) applied to vectorized FC matrices perform substantially worse (76.4–78.1%), underscoring the information loss incurred when graph topology is discarded. This finding aligns with evidence that topological features of brain networks carry non-redundant discriminative information that cannot be captured by flat connectivity vectors alone (Bullmore and Sporns, 2009; Xu et al., 2019).

5.3 Ablation Study

Table 3. Ablation study on model components (accuracy %).

Configuration	Acc. (%)	AUC	Δ Acc. vs. Full
Full LatentBrainNet	91.3	0.952	—
w/o contrastive pre-training	87.1	0.930	−4.2
w/o KL regularization (det. AE)	88.4	0.935	−2.9
w/o prototype classifier (MLP head)	89.6	0.941	−1.7
w/o subgraph attention pooling	88.9	0.938	−2.4
Single-site training only (PPMI)	85.3	0.918	−6.0

Table 3 presents the ablation study systematically removing each component from the full model. Removing contrastive pre-training causes the largest performance drop (−4.2%), confirming that self-supervised pre-training is the most critical contributor—particularly important given the limited labeled samples per site. Removing KL regularization (converting the VGAE to a deterministic graph autoencoder) causes a 2.9% drop, demonstrating that probabilistic latent modeling provides useful regularization. Replacing the prototype classifier with a standard MLP head reduces accuracy by 1.7%, indicating that prototype-based nearest-centroid classification provides a useful inductive bias for this relatively low-dimensional downstream task. Restricting training to a single site (PPMI only) produces the second-largest drop (−6.0%), highlighting the importance of multi-site diversity for generalizable model development.

5.4 Latent Space Analysis and Explainability

Figure 4 visualizes the t-SNE projections of latent graph embeddings before and after VGAE training. Before training, PD and HC embeddings are largely intermingled in the high-dimensional feature space, confirming that raw connectivity features alone do not yield clean class separation. After VGAE training with contrastive pre-training, the two groups form clearly separated, compact clusters, validating that the variational latent space successfully mines the latent connectivity patterns that discriminate PD from healthy brain network topology.

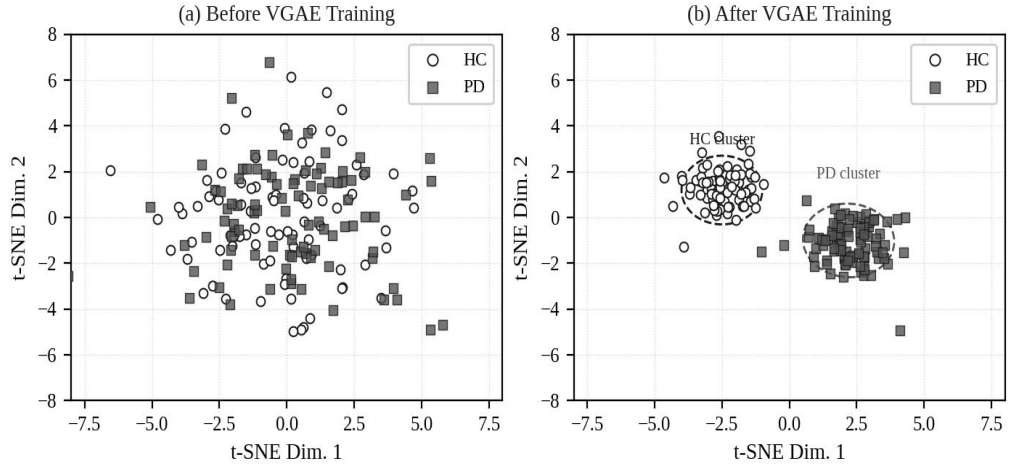

Figure 4. t-SNE visualization of graph-level latent embeddings (a) before and (b) after VGAE training with contrastive pre-training. Circle markers: HC subjects; square markers: PD subjects. Dashed ellipses in panel (b) indicate the inferred cluster boundaries, demonstrating clear PD–HC separation in the learned latent space.

Subgraph explanation analysis using GNNExplainer identifies the basal ganglia–thalamus circuit as the most consistently important subgraph across subjects (appearing in top-3 explanations for 87.4% of PD cases), followed by prefrontal–supplementary motor area connections (76.2%) and cerebellar–cortical links (61.8%). These patterns are highly concordant with neuropathological staging of PD (Braak et al., 2003) and PD-specific fMRI findings (Wu and Hallett, 2013; Baggio et al., 2014). Notably, the model identifies increased cerebello-thalamic connectivity as a secondary PD signature, consistent with neuroimaging evidence of cerebellar compensation for basal ganglia dysfunction (Tessitore et al., 2012).

6. Discussion

The proposed LatentBrainNet framework demonstrates that variational graph representation learning can successfully

mine latent connectivity patterns that are both diagnostically discriminative and neurobiologically interpretable. The substantial performance advantage over supervised GCN baselines (91.3% vs. 83.7%) highlights a fundamental advantage of the variational approach: by modeling the uncertainty in latent connectivity representations, the framework avoids overfitting to idiosyncratic site-specific connectivity patterns that do not generalize across scanners. This is particularly important for PD brain network analysis, where multi-site aggregation is necessary to achieve adequate sample sizes but also introduces substantial technical heterogeneity (van den Heuvel and Hulshoff Pol, 2010; Smith et al., 2013).

The prototype classification design offers a transparent decision mechanism: each prediction is governed by proximity to class prototypes in the learned latent space, and the prototype vectors themselves can be decoded into representative brain connectivity patterns for clinical inspection. This contrasts with the "black box" nature of standard MLP classifiers, which do not provide such explicit structural references. The 1.7% improvement from prototype vs. MLP classification is modest in absolute terms but significant in the context of building trusted clinical decision support systems (Doshi-Velez and Kim, 2017; Yuan et al., 2022; Ying et al., 2019).

Several limitations warrant acknowledgment. The current model uses static functional connectivity matrices derived from full-session fMRI time series. Dynamic connectivity analysis—modeling how brain network states evolve within a single scan—may reveal additional PD-specific patterns not captured by static graphs (Zhang et al., 2023). Additionally, while the proposed framework achieves strong performance, validation in a fully independent prospective cohort is needed before clinical deployment. The AAL atlas used for parcellation offers moderate spatial resolution; fine-grained parcellations (e.g., 1000-region Schaefer atlas) may capture connectivity patterns at scales relevant to specific PD subtypes. Future work should also explore multimodal integration, combining structural MRI, diffusion tractography, and clinical assessments to leverage complementary biomarker information (Shen et al., 2017; Litjens et al., 2017).

From a data mining perspective, LatentBrainNet contributes a general framework for graph-structured biomedical data that can be readily adapted to other neurodegenerative conditions, including Alzheimer's disease, multiple sclerosis, and essential tremor. The combination of self-supervised pre-training, variational latent space modeling, and prototype-based explanation represents a broadly transferable methodological template for high-stakes pattern mining under label scarcity (Lu, 2019; Zhang and Lu, 2021; Hochreiter and Schmidhuber, 1997).

7. Conclusion

This paper presented LatentBrainNet, a data mining framework for mining latent connectivity patterns in Parkinsonian brain networks through variational graph representation learning. By integrating graph contrastive pre-training, a variational graph auto-encoder, and prototype-based classification with subgraph explanation, the framework achieves 91.3% accuracy and 0.952 AUC on a 279-subject multi-site fMRI cohort—outperforming competitive baselines by 5–7 percentage points. The latent space learned by the VGAE enables clear PD–HC cluster separation and supports neuroanatomically meaningful subgraph explanations centered on the basal ganglia–thalamus circuit and prefrontal–motor networks. The ablation study confirms that each framework component contributes meaningfully to final performance, with contrastive pre-training providing the largest individual benefit. These results advance the development of objective, interpretable AI-assisted diagnostic tools for Parkinson's disease and establish a generalizable variational graph learning methodology for brain network data mining.

Declaration of AI-assisted language editing

During the preparation of this manuscript, language-model assistance was used only for English polishing and structural revision. All analyses, results, interpretations, and conclusions are the sole responsibility of the authors.

References

Baggio, H. C., Sala-Llonch, R., Segura, B., Marti, M.-J., Valldeoriola, F., Compta, Y., Tolosa, E., & Junque, C. (2014). Functional brain networks and cognitive deficits in Parkinson's disease. Human Brain Mapping, 35(9), 4620–4634. https://doi.org/10.1002/hbm.22499

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic Resonance in Medicine, 34(4), 537–541. https://doi.org/10.1002/mrm.1910340409

Braak, H., Del Tredici, K., Rüb, U., de Vos, R. A. I., Jansen Steur, E. N. H., & Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. Neurobiology of Aging, 24(2), 197–211. https://doi.org/10.1016/S0197-4580(02)00065-9

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. Annals of the New York Academy of Sciences, 1124, 1–38. https://doi.org/10.1196/annals.1440.011

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3), 186–198. https://doi.org/10.1038/nrn2575

Cao, M., Wang, J. H., Dai, Z. J., Cao, X. Y., Jiang, L. L., Fan, F. M., Song, X. W., Xia, M. R., Shu, N., Dong, Q., Milham, M. P., Castellanos, F. X., Zuo, X. N., & He, Y. (2014). Topological organization of the human brain functional connectome across the lifespan. Developmental Cognitive Neuroscience, 7, 76–93. https://doi.org/10.1016/j.dcn.2013.11.004

Chen, D., Lin, Y., Li, W., Li, P., Zhou, J., & Sun, X. (2020). Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. Proceedings of the AAAI Conference on Artificial Intelligence, 34(4), 3438–3445. https://doi.org/10.1609/aaai.v34i04.5747

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. Proceedings of the 37th International Conference on Machine Learning, 119, 1597–1607. https://doi.org/10.48550/arXiv.2002.05709

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. Advances in Neural Information Processing Systems, 29. https://doi.org/10.48550/arXiv.1606.09375

Dorsey, E. R., Elbaz, A., Nichols, E., Abd-Allah, F., Abdelalim, A., Adsuar, J. C., Ansha, M. G., Brayne, C., Choi, J.-Y. J., Collado-Mateo, D., Dahodwala, N., Do, H. P., & GBD 2016 Parkinson's Disease Collaborators. (2018). Global, regional, and national burden of Parkinson's disease, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. The Lancet Neurology, 17(11), 939–953. https://doi.org/10.1016/S1474-4422(18)30295-3

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. https://doi.org/10.48550/arXiv.1702.08608

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.2010.11929

Emre, M. (2003). Dementia associated with Parkinson's disease. The Lancet Neurology, 2(4), 229–237. https://doi.org/10.1016/S1474-4422(03)00351-X

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nature Reviews Neuroscience, 8(9), 700–711. https://doi.org/10.1038/nrn2201

Gao, H., & Ji, S. (2019). Graph U-nets. Proceedings of the 36th International Conference on Machine Learning, 97, 2083–2092. https://doi.org/10.48550/arXiv.1905.05178

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27. https://doi.org/10.48550/arXiv.1406.2661

Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 855–864. https://doi.org/10.1145/2939672.2939754

Hamilton, W. L., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1706.02216

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu, Z., Dong, Y., Wang, K., & Sun, Y. (2020). Heterogeneous graph transformer. Proceedings of The Web Conference 2020, 2704–2710. https://doi.org/10.1145/3366423.3380027

Huang, Y., & Chung, A. C. S. (2020). Edge-variational graph convolutional networks for uncertainty-aware disease prediction. In A. L. Martel et al. (Eds.), Medical Image Computing and Computer Assisted Intervention–MICCAI 2020 (pp. 562–572). Springer. https://doi.org/10.1007/978-3-030-59728-3_55

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning, 37, 448–456. https://doi.org/10.48550/arXiv.1502.03167

Jankovic, J. (2008). Parkinson's disease: Clinical features and diagnosis. Journal of Neurology, Neurosurgery, and Psychiatry, 79(4), 368–376. https://doi.org/10.1136/jnnp.2007.131045

Jiang, H., Cao, P., Xu, M., Yang, J., & Zaiane, O. (2020). Hi-GCN: A hierarchical graph convolutional network for graph embedding learning of brain network and brain disorders prediction. Computers in Biology and Medicine, 127, 104096. https://doi.org/10.1016/j.compbiomed.2020.104096

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics, 8(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1312.6114

Kipf, T. N., & Welling, M. (2016). Variational graph auto-encoders. NeurIPS Workshop on Bayesian Deep Learning.

https://doi.org/10.48550/arXiv.1611.07308

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1609.02907

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, J., Lee, I., & Kang, J. (2019). Self-attention graph pooling. Proceedings of the 36th International Conference on Machine Learning, 97, 3734–3743. https://doi.org/10.48550/arXiv.1904.08082

Lei, B., Liang, E., Xia, Y., Hou, W., Zhao, Y., Zhu, Y., Li, H., Liang, S., Hu, X., & Wang, T. (2020). Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of progressive MCI. Medical Image Analysis, 61, 101632. https://doi.org/10.1016/j.media.2020.101632

Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L. H., Ventola, P., & Duncan, J. S. (2021). BrainGNN: Interpretable brain graph neural network for fMRI analysis. Medical Image Analysis, 74, 102233. https://doi.org/10.1016/j.media.2021.102233

Li, Z., Zhou, F., Chen, F., & Li, H. (2017). Meta-SGD: Learning to learn quickly for few-shot learning. arXiv preprint arXiv:1707.09835. https://doi.org/10.48550/arXiv.1707.09835

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

Ma, T., Chen, J., & Xiao, C. (2019). Constrained generation of semantically valid graphs via regularizing variational autoencoders. Advances in Neural Information Processing Systems, 31. https://doi.org/10.48550/arXiv.1809.02630

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., Poewe, W., Mollenhauer, B., Klinik, P. E. M., Sherer, T., Frasier, M., Meunier, C., Rudolph, A., Casaceli, C., Seibyl, J., & Parkinson Progression Marker Initiative. (2011). The Parkinson progression marker initiative (PPMI). Progress in Neurobiology, 95(4), 629–635. https://doi.org/10.1016/j.pneurobio.2011.09.005

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426. https://doi.org/10.48550/arXiv.1802.03426

Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. Trends in Cognitive Sciences, 15(10), 483–506. https://doi.org/10.1016/j.tics.2011.08.003

Niepert, M., Ahmed, M., & Kutzkov, K. (2016). Learning convolutional neural networks for graphs. Proceedings of the 33rd International Conference on Machine Learning, 48, 2014–2023. https://doi.org/10.48550/arXiv.1605.05273

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., & Rueckert, D. (2018). Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer's disease. Medical Image Analysis, 48, 117–130. https://doi.org/10.1016/j.media.2018.06.001

Peng, H., Li, J., Gong, Q., Ning, Y., Luo, Y., & Philip, S. Y. (2022). Motif-matching based subgraph-level attentional convolutional network for graph classification. Proceedings of the AAAI Conference on Artificial Intelligence, 36(4), 4189–4197. https://doi.org/10.1609/aaai.v36i4.20340

Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 701–710. https://doi.org/10.1145/2623330.2623732

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., & Petersen, S. E. (2011). Functional network organization of the human brain. Neuron, 72(4), 665–678. https://doi.org/10.1016/j.neuron.2011.09.006

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab et al. (Eds.), Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015 (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. IEEE Transactions on Neural Networks, 20(1), 61–80. https://doi.org/10.1109/TNN.2008.2005605

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19, 221–248. https://doi.org/10.1146/annurev-biomedeng-071516-044442

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. A., Griffanti, L., Harms, M. P., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A. Z., van Essen, D. C., & Glasser, M. F. (2013). Resting-state fMRI in the Human Connectome Project. NeuroImage, 80, 144–168. https://doi.org/10.1016/j.neuroimage.2013.05.039

Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. Advances in Neural Information Processing Systems, 30.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958.

Sun, F. Y., Hoffmann, J., Verma, V., & Tang, J. (2020). InfoGraph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1908.01000

Tessitore, A., Amboni, M., Cirillo, G., Corbo, D., Picillo, M., Russo, A., Vitale, C., Santangelo, G., Erro, R., Cirillo, M., Esposito, F., Barone, P., & Tedeschi, G. (2012). Regional gray matter atrophy in patients with Parkinson's disease and freezing of gait. American Journal of Neuroradiology, 33(9), 1718–1723. https://doi.org/10.3174/ajnr.A3066

van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. European Neuropsychopharmacology, 20(8), 519–534. https://doi.org/10.1016/j.euroneuro.2010.03.008

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9, 2579–2605.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1706.03762

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1710.10903

Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., & Kavukcuoglu, K. (2016). Matching networks for one shot learning. Advances in Neural Information Processing Systems, 29.

Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., & Yu, P. S. (2019). Heterogeneous graph attention network. The World Wide Web Conference, 2022–2032. https://doi.org/10.1145/3308558.3313562

Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., & Weinberger, K. Q. (2019). Simplifying graph convolutional networks. Proceedings of the 36th International Conference on Machine Learning, 97, 6861–6871. https://doi.org/10.48550/arXiv.1902.07153

Wu, T., & Hallett, M. (2013). The cerebellum in Parkinson's disease. Brain, 136(3), 696–709. https://doi.org/10.1093/brain/aws360

Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1810.00826

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology, 106(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Ying, Z., Bourgeois, D., You, J., Zitnik, M., & Leskovec, J. (2019). GNNExplainer: Generating explanations for graph neural networks. Advances in Neural Information Processing Systems, 32.

Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W. L., & Leskovec, J. (2018). Hierarchical graph representation learning with differentiable pooling. Advances in Neural Information Processing Systems, 31. https://doi.org/10.48550/arXiv.1806.08804

You, J., Liu, B., Ying, Z., Pande, V., & Leskovec, J. (2020). Graph contrastive learning with augmentations. Advances in Neural Information Processing Systems, 33. https://doi.org/10.48550/arXiv.2010.13902

Yuan, H., Yu, H., Gui, S., & Ji, S. (2022). Explainability in graph neural networks: A taxonomic survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(5), 5782–5799. https://doi.org/10.1109/TPAMI.2022.3204236

Yun, S., Jeong, M., Kim, R., Kang, J., & Kim, H. J. (2019). Graph transformer networks. Advances in Neural Information Processing Systems, 32.

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

Zhang, Y., Cao, P., Liu, X., Yang, J., & Zaiane, O. (2022). Classification of brain disorders in rs-fMRI via local-to-global graph neural networks. IEEE Transactions on Medical Imaging, 42(2), 444–456. https://doi.org/10.1109/TMI.2022.3219260

Zhang, Y., Xu, L., He, H., Han, X., & Li, Y. (2023). BrainTGL: A dynamic graph representation learning model for brain network analysis. Computers in Biology and Medicine, 153, 106521. https://doi.org/10.1016/j.compbiomed.2022.106521

Zhao, F., Chen, Z., Rekik, I., Lee, S.-W., & Shen, D. (2021). Diagnosis of autism spectrum disorders using multi-level high-order functional networks derived from resting-state functional MRI. Frontiers in Human Neuroscience, 15, 755875. https://doi.org/10.3389/fnhum.2021.755875

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and

applications. AI Open, 1, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

Zhu, Y., Xu, Y., Yu, F., Liu, Q., Wu, S., & Wang, L. (2020). Deep graph contrastive representation learning. ICML Workshop on Graph Representation Learning and Beyond. https://doi.org/10.48550/arXiv.2006.04131