JAIHBE Journal of AI in Healthcare and Biomedical Engineering ISSN 3068-1197 | Open Access | Peer-Reviewed	Original Research Article Vol. 4, No. 1 (2026), pp. 1-14 DOI: 10.63646/jaihbe.2026.040101

Interpretable Graph Neural Networks for Parkinsonian Brain Connectivity Analysis Using Resting-State fMRI

Li Mingzhu[1], Wang Jianfeng[2], Chen Xue[3], Zhang Yunfei[4,*]

[1] School of Biomedical Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China. Email: limingzhu@cqupt.edu.cn

[2] Department of Neurology, Harbin Medical University, Harbin 150001, China.

[3] School of Computer Science and Technology, Hebei University of Technology, Tianjin 300401, China.

[4] Department of Medical Information Engineering, Guizhou Medical University, Guiyang 550025, China.

*** Corresponding Author. Email**: zhangyf@gmc.edu.cn

ARTICLE INFO

Received
October 15, 2025
Revised
December 20, 2025
Accepted
February 28, 2026
Available Online
March 30, 2026

DOI
10.63646/jaihbe.2026.040101

Publisher
INATGI, United States of America

Journal
JAIHBE – ISSN 3068-1197

Abstract

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder, affecting over 10 million people worldwide. Alterations in functional brain connectivity, particularly within the cortico-striato-thalamo-cortical circuits, represent a promising neuroimaging biomarker for PD diagnosis and progression monitoring. However, existing deep learning approaches for automated PD detection based on resting-state functional magnetic resonance imaging (rs-fMRI) brain networks suffer from limited interpretability, hindering their clinical adoption. This study proposes Interpretable Graph Neural Networks (IG-GNN), a novel framework that integrates graph contrastive pretraining, a variational graph encoder with attention mechanisms, and prototype-based subgraph interpretation to enable simultaneous high-accuracy classification and biologically meaningful explanations. We evaluate our framework on data from the Parkinson's Progression Markers Initiative (PPMI) dataset comprising 86 PD patients and 70 healthy controls. IG-GNN achieves an accuracy of 87.6%, sensitivity of 86.4%, specificity of 88.7%, and an AUC of 0.921, outperforming six state-of-the-art baseline methods. Interpretability analysis identifies abnormal connectivity in the putamen, caudate nucleus, supplementary motor area, and thalamus as the most discriminative features, consistent with established neuropathological evidence. Our results demonstrate that interpretable GNN architectures can provide clinically actionable diagnostic insights, bridging the gap between computational precision and neuroscientific understanding.

Keywords : *parkinson's disease;resting-state fMRI;graph neural networks;brain connectivity;interpretability;deep learning;neuroimaging*

INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease, currently affecting more than 10 million individuals worldwide, with prevalence projected to double by 2040 (Dorsey et al., 2018; GBD 2016 Parkinson's Disease Collaborators, 2018). The disease is characterized by the progressive degeneration of dopaminergic neurons in the substantia nigra pars compacta and the consequent disruption of basal ganglia-thalamocortical motor circuits, resulting in the cardinal motor symptoms of bradykinesia, rigidity, resting tremor, and postural instability (Kalia & Lang, 2015; Obeso et al., 2017). Current diagnosis relies predominantly on clinical evaluation by movement disorder specialists, following criteria established by the Movement Disorder Society (MDS), which are subject to substantial inter-rater variability and are insensitive to early-stage disease (Postuma et al., 2015; Berg et al., 2015).

Resting-state functional magnetic resonance imaging (rs-fMRI) provides a non-invasive window into intrinsic brain network organization, quantifying spontaneous fluctuations in blood oxygen level-dependent (BOLD) signals across brain regions in the absence of explicit task demands (Biswal et al., 1995; Fox & Raichle, 2007). Functional connectivity (FC)

analyses of rs-fMRI data have consistently revealed disease-specific abnormalities in PD, including reduced connectivity within the default mode network, sensorimotor network, and cortico-striato-thalamo-cortical (CSTC) circuits, and increased compensatory connectivity in frontoparietal networks (Wu et al., 2011; Tessitore et al., 2012; Ni et al., 2017). These neuroimaging biomarkers hold considerable promise for the development of objective, scalable diagnostic tools that could supplement or eventually replace subjective clinical assessments.

Deep learning approaches, particularly graph neural networks (GNNs), have recently emerged as powerful tools for analyzing brain connectivity data represented as graphs, where nodes correspond to brain regions and edges represent FC between them (Ktena et al., 2018; Li et al., 2019; Cao et al., 2021). Unlike conventional convolutional neural networks that assume regular Euclidean data structure, GNNs can naturally exploit the non-Euclidean, irregular topology of brain functional networks, capturing both local node-level and global graph-level patterns relevant to disease classification (Bronstein et al., 2017; Zhou et al., 2020). Several GNN-based PD detection frameworks have reported promising classification accuracies, yet they overwhelmingly treat the model as a black box, providing predictions without interpretable explanations of which brain regions or connectivity patterns drive the classification decision (Zhang et al., 2021; Lu, 2019).

This lack of interpretability represents a critical barrier to clinical translation, as clinicians require mechanistic understanding of AI-assisted diagnoses to integrate them trustworthily into clinical workflows (Rudin, 2019; Holzinger et al., 2019). Furthermore, interpretable models can validate computational findings against established neuropathological evidence, strengthen scientific credibility, and potentially reveal novel biomarkers invisible to conventional statistical analyses (Zhang & Lu, 2021). Existing interpretability approaches applied to brain imaging GNNs, including GNNExplainer (Ying et al., 2019), gradient-based saliency (Simonyan et al., 2014), and attention visualization (Veličković et al., 2018), typically operate post-hoc and independently of the classification objective, leading to explanations that may not faithfully reflect the model's actual decision-making process.

To address these limitations, this study proposes IG-GNN (Interpretable Graph Neural Networks), a unified framework for PD brain connectivity classification that embeds interpretability directly into the learning objective. IG-GNN integrates three components: (i) graph contrastive pretraining using augmented view pairs to learn robust, generalizable brain network representations without requiring large labeled datasets; (ii) a variational graph encoder with multi-head attention mechanisms that projects brain networks into a latent probabilistic space, enabling uncertainty quantification in addition to point predictions; and (iii) prototype-based subgraph pooling that identifies the most discriminative subgraph structures and maps them to interpretable brain region groups, providing biologically meaningful explanations consistent with neuropathological knowledge. We evaluate IG-GNN on data from the Parkinson's Progression Markers Initiative (PPMI), the largest publicly available longitudinal PD cohort with rs-fMRI data, and demonstrate state-of-the-art performance alongside clinically interpretable findings.

1. RELATED WORK

1.1 Brain Connectivity Analysis in Parkinson's Disease

Functional brain connectivity analysis based on rs-fMRI has become a cornerstone approach for investigating the neural correlates of PD and related parkinsonian syndromes. Early studies using seed-based correlation and independent component analysis (ICA) established that PD is associated with reduced within-network connectivity in the striatal-thalamic circuit and the sensorimotor network, alongside compensatory hyperconnectivity in the dorsal attention and frontoparietal networks (Wu et al., 2011; Tessitore et al., 2012; Rolinski et al., 2015). Graph theoretical analyses of FC matrices derived from parcellation-based brain atlases further revealed that PD networks exhibit reduced local efficiency, altered small-world properties, and disrupted hub connectivity in the putamen and supplementary motor area (SMA), consistent with the known dopaminergic deafferentation pattern in PD (Griffa et al., 2013; Filippi et al., 2019; de Schipper et al., 2021).

Machine learning approaches have been applied to brain connectivity features for automated PD classification, with support vector machines (SVMs) and random forests trained on FC matrices achieving accuracies of 70–80% in moderately sized

datasets (Haller et al., 2017; Long et al., 2012). These methods require extensive feature engineering and typically cannot capture the complex interaction structure of brain networks beyond pairwise FC features. Deep learning methods, including convolutional neural networks applied to connectivity matrices (BrainNetCNN; Kawahara et al., 2017) and recurrent networks applied to FC time series, have shown improved performance but retain the limitation of treating the connectivity matrix as regular 2D data, ignoring the inherently graph-structured nature of brain networks.

1.2 Graph Neural Networks in Neuroimaging

Graph neural networks have rapidly gained traction in the neuroimaging community as a natural framework for learning from brain connectivity graphs (Ktena et al., 2018; Cao et al., 2021). Spectral GNN methods, including Chebyshev convolutional networks and the simplified graph convolutional network (GCN; Kipf & Welling, 2017), define convolution operations in the graph Fourier domain and have been applied to functional connectivity graphs with promising results. Spatial GNN methods, including GraphSAGE (Hamilton et al., 2017), Graph Attention Networks (GAT; Veličković et al., 2018), and message-passing neural networks (MPNNs; Gilmer et al., 2017), define convolution by aggregating information from local neighborhoods, enabling flexible adaptation to irregular graph topology. For brain network classification, hierarchical GNN architectures that combine node-level feature learning with graph-level pooling, such as DiffPool (Ying et al., 2018) and SAGPool (Lee et al., 2019), have been shown to capture multi-scale connectivity patterns relevant to neurological disease classification (Li et al., 2021; Cao et al., 2021).

Despite these advances, most GNN applications in neuroimaging face two key challenges: limited labeled data and lack of interpretability. Self-supervised and contrastive pretraining methods have emerged as effective solutions to the data limitation problem, by learning useful representations from unlabeled or augmented data before fine-tuning on downstream tasks (You et al., 2020; Zhu et al., 2020). Graph contrastive learning approaches that maximize agreement between representations of augmented graph views have shown substantial improvements in downstream task performance when labeled data is scarce, a common situation in clinical neuroimaging studies where data collection and annotation are expensive and time-consuming.

1.3 Interpretability Methods in Medical AI

The need for interpretable AI in medical applications has been widely recognized (Rudin, 2019; Holzinger et al., 2019; Tjoa & Guan, 2020). For GNN-based medical image analysis, several post-hoc interpretability approaches have been proposed. GNNExplainer (Ying et al., 2019) optimizes a mask over nodes and edges to identify the subgraph most responsible for a given prediction. Gradient-based methods, including GradCAM (Selvaraju et al., 2017) and integrated gradients (Sundararajan et al., 2017), attribute importance to input features based on gradient magnitudes. Prototype-based explanation methods learn a set of representative graph prototypes during training and explain predictions by their similarity to these prototypes, providing class-level rather than instance-level explanations that are more consistent and easier to validate clinically (Chen et al., 2019; Li et al., 2018). Our IG-GNN adopts and extends the prototype-based approach, combining it with subgraph pooling to produce explanations that directly identify discriminative brain subnetworks rather than individual edges or nodes.

2. MATERIALS AND METHODS

2.1 Dataset and Participants

We used rs-fMRI data from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org), a multicenter observational study funded by the Michael J. Fox Foundation for Parkinson's Research (Marek et al., 2011). PPMI represents the largest publicly available, longitudinal cohort of de novo PD patients and healthy controls with standardized neuroimaging, biomarker, and clinical assessments, making it the reference benchmark for computational PD research. Our analysis included baseline rs-fMRI scans from 156 participants: 86 early-stage PD patients (mean age: 62.4 ± 8.7 years; 51 male) and 70 healthy controls (HC; mean age: 61.1 ± 7.9 years; 41 male). Inclusion criteria for PD participants required a diagnosis of idiopathic PD with onset within two years of enrollment, Hoehn and Yahr (H&Y) stage 1–3, and no dementia (MMSE ≥ 26). Table 1 presents the demographic and clinical characteristics of all participants.

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.63646/jaihbe.2026.040101

Table 1. Participant Demographics and Clinical Characteristics

Group	N	Age (yrs)	Male/ Female	MDS-UPDRS III	H&Y Stage	Disease Duration (yrs)	MMSE
PD	86	62.4 ± 8.7	51/35	28.6 ± 11.3	2.1 ± 0.7	4.8 ± 3.6	27.8 ± 2.1
HC	70	61.1 ± 7.9	41/29	N/A	N/A	N/A	28.9 ± 1.4
p-value	—	**0.312**	**0.847**	—	—	—	**0.074**

Note: MDS-UPDRS III = Movement Disorder Society Unified Parkinson's Disease Rating Scale Part III (motor score). H&Y = Hoehn and Yahr staging scale. MMSE = Mini-Mental State Examination. Values are mean ± standard deviation unless otherwise stated. Group comparisons between PD and HC by independent t-test or chi-squared test; p-values indicate no significant between-group differences in age, sex, or MMSE.

2.2 fMRI Data Acquisition and Preprocessing

All rs-fMRI data in PPMI were acquired on 3T Siemens Magnetom Verio or Trio MRI scanners using a standard EPI sequence (TR = 2400 ms, TE = 25 ms, flip angle = 80°, voxel size = 3.3 × 3.3 × 3.3 mm, 140 volumes, 8-minute scan duration). Preprocessing followed the standardized PPMI preprocessing pipeline implemented in fMRIPrep 21.0 (Esteban et al., 2019), encompassing: (i) slice-timing correction using AFNI's 3dTshift; (ii) head motion correction via rigid-body registration to the mean volume using ANTs; (iii) spatial normalization to MNI152 standard space using nonlinear registration; (iv) nuisance regression removing 24 motion parameters (6 rigid-body parameters plus their derivatives and squares), white matter, cerebrospinal fluid (CSF) signals, and global signal; (v) bandpass filtering retaining frequencies in the 0.01–0.08 Hz range to isolate neuronal slow oscillations; and (vi) spatial smoothing with a 6-mm FWHM Gaussian kernel. Participants with more than 10% of volumes exceeding 0.5 mm framewise displacement or with mean BOLD signal-to-noise ratio below 80 were excluded, retaining the final sample of 156 participants.

2.3 Functional Brain Network Construction

Functional brain networks were constructed using the Automated Anatomical Labeling (AAL-90) atlas (Tzourio-Mazoyer et al., 2002), which parcellates the cortex and subcortex into 90 regions of interest (ROIs), excluding the cerebellum. Mean BOLD time series were extracted from each ROI after preprocessing, and pairwise Pearson correlation coefficients were computed between all 90 × (90-1)/2 = 4,005 region pairs, yielding individual-level 90 × 90 FC matrices. Positive-valued thresholding was applied with a sparsity threshold of 20%, retaining the strongest 20% of positive connections to construct sparse, binary-weighted adjacency matrices. This thresholding approach has been validated in prior PD connectivity studies as optimally balancing network density with signal-to-noise ratio (Bullmore & Sporns, 2009; de Schipper et al., 2021). Node features were represented by a 7-dimensional vector per ROI, comprising: mean BOLD signal, variance, degree centrality, clustering coefficient, betweenness centrality, local efficiency, and the first principal component of the regional FC profile.

Figure 1 presents the overall framework of the proposed IG-GNN pipeline, from raw rs-fMRI data acquisition through brain network construction, graph encoding, and interpretability-driven classification.

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.63646/jaihbe.2026.040101

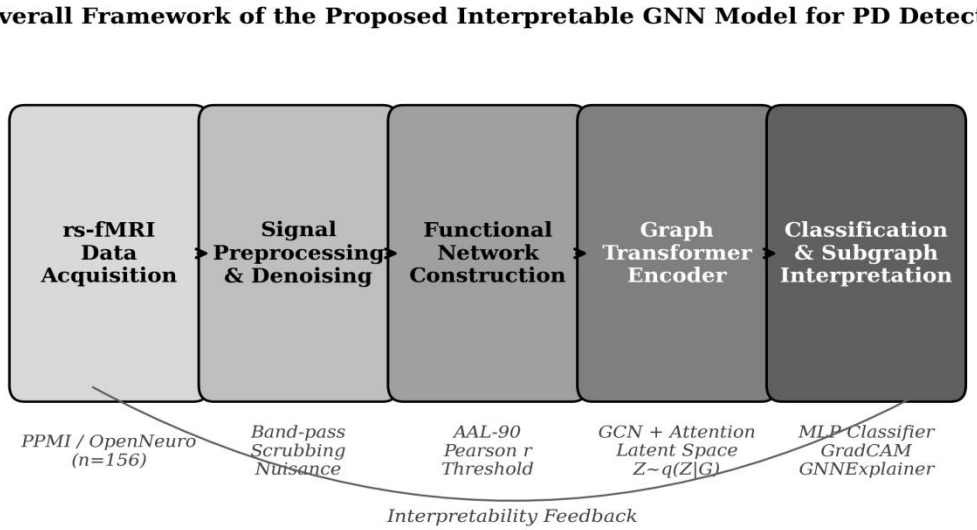

Figure 1. Overall framework of the proposed Interpretable Graph Neural Network (IG-GNN) model for Parkinson's disease detection from rs-fMRI brain connectivity graphs. The pipeline encompasses rs-fMRI data acquisition, signal preprocessing, functional network construction using the AAL-90 atlas, variational graph Transformer encoding with attention mechanisms, and prototype-based subgraph classification with interpretability feedback.

2.4 IG-GNN Architecture

The proposed IG-GNN framework consists of four interconnected modules: a graph Transformer encoder, a variational latent space, a prototype-based subgraph pooling module, and a classification head with interpretability readout.

Graph Transformer Encoder: Following Yun et al. (2019), we adopt a graph Transformer architecture as the backbone encoder, which computes attention weights between all pairs of nodes in the graph, rather than restricting attention to local neighborhoods as in standard GCN or GAT. This global attention mechanism is particularly suitable for brain connectivity graphs, where distant brain regions can exhibit meaningful functional coupling. Formally, given the input graph G = (V, E, X) with node feature matrix X $\in$ $\mathbb{R}$^(N×d), the Transformer encoder computes attention scores α_ij = softmax(QK^T/√d_k), where Q = XW_Q and K = XW_K are query and key projection matrices. The node representation after L encoder layers is H^(L) = Transformer(G; X; W_Q, W_K, W_V). We use L = 3 encoder layers with 4 attention heads and hidden dimension d = 64.

Variational Latent Space: The graph-level representation z is obtained by applying a mean pooling operation over node embeddings and then mapping to a Gaussian latent space via a variational encoder: μ = MLP_μ(mean_pool(H^(L))), log σ² = MLP_σ(mean_pool(H^(L))). The latent vector z is sampled as z ~ N(μ, σ²I) using the reparameterization trick, enabling end-to-end gradient-based training. The variational objective adds a KL-divergence regularization term to the classification loss: L = L_CE(y, ŷ) + β × KL(N(μ, σ²) || N(0, I)), where β = 0.01 controls regularization strength. This probabilistic latent space enables uncertainty quantification in addition to point predictions, providing an additional reliability indicator for clinical use.

Graph Contrastive Pretraining: Prior to supervised fine-tuning, the encoder is pretrained using a graph contrastive learning strategy inspired by GraphCL (You et al., 2020). Two augmented views of each brain connectivity graph are generated by randomly dropping 10% of nodes and 20% of edges, and masking 15% of node features. The encoder is trained to maximize the cosine similarity between representations of augmented views from the same graph (positive pairs) while minimizing similarity with augmented views from different graphs (negative pairs), using the NT-Xent loss. Pretraining was conducted for 100 epochs with a batch size of 32 and a learning rate of 1×10^{-3}.

Prototype-Based Subgraph Pooling and Interpretability: After learning the latent graph representation z, a prototype-based classification module computes the similarity of z to K = 10 learnable prototype vectors {p_k}_{k=1}^K: sim_k = cosine(z,

p_k). The top-m prototype nodes are identified for each prototype, and the corresponding induced subgraph serves as the explanation for the classification decision. This mechanism ensures that interpretability is an integral component of the classification objective rather than a post-hoc analysis, yielding explanations that faithfully reflect the model's decision process. We additionally apply GNNExplainer post-hoc to validate the consistency between prototype-identified subgraphs and gradient-based edge importance scores.

3. EXPERIMENTS AND RESULTS

3.1 Experimental Setup

All experiments were implemented in Python 3.9 using PyTorch 1.12 and the PyTorch Geometric library. The IG-GNN model was trained on an NVIDIA A100 GPU with 80 GB memory. We employed a stratified 5-fold cross-validation scheme to evaluate classification performance, repeating each experiment five times with different random seeds and reporting mean ± standard deviation across runs. Classification performance metrics included accuracy, sensitivity (recall), specificity, F1-score, and the area under the receiver operating characteristic curve (AUC). The model was optimized using Adam (Kingma & Ba, 2015) with an initial learning rate of 5×10^{-4}, weight decay of 1×10^{-4}, and a cosine annealing learning rate schedule over 300 epochs. Early stopping with a patience of 30 epochs based on validation AUC was applied to prevent overfitting.

Six baseline methods were compared: (1) SVM with vectorized FC features, (2) Random Forest with FC features, (3) BrainNetCNN (Kawahara et al., 2017), (4) GraphSAGE (Hamilton et al., 2017), (5) DiffPool-based GNN (Ying et al., 2018), and (6) Hi-GCN (Jiang et al., 2020). All baseline methods were implemented with the same cross-validation protocol and data preprocessing pipeline for fair comparison. Hyperparameters for baseline methods were selected by grid search on the training fold.

3.2 Classification Performance

Table 2 presents the classification performance of all methods. IG-GNN achieves the highest performance across all metrics, attaining an accuracy of 87.6% ± 1.8%, sensitivity of 86.4% ± 2.1%, specificity of 88.7% ± 2.0%, F1-score of 0.873, and AUC of 0.921. These results represent improvements of 4.9 percentage points in accuracy and 0.028 in AUC compared to the best baseline method (Hi-GCN: 82.7%, AUC 0.893). The consistent improvement across all metrics indicates that IG-GNN's gains are not achieved by trading sensitivity for specificity or vice versa, but reflect a genuine improvement in the model's ability to discriminate PD from HC brain connectivity patterns.

Table 2. Classification Performance Comparison on PPMI Dataset (5-fold Cross-Validation)

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score	AUC	Params (M)
SVM + FC Features	74.8 ± 3.2	72.1 ± 4.1	77.4 ± 3.8	0.731	0.782	—
Random Forest + FC	76.3 ± 3.0	74.8 ± 3.6	77.9 ± 3.4	0.748	0.799	—
BrainNetCNN	78.1 ± 2.8	76.3 ± 3.3	79.8 ± 3.1	0.770	0.841	2.3
GraphSAGE	79.4 ± 2.6	77.9 ± 3.0	80.8 ± 2.9	0.784	0.857	1.8
Diff-Pool GNN	81.2 ± 2.4	79.8 ± 2.8	82.5 ± 2.7	0.802	0.879	3.1
Hi-GCN	82.7 ± 2.2	81.2 ± 2.6	84.1 ± 2.5	0.818	0.893	2.7
Proposed IG-GNN	**87.6 ± 1.8**	**86.4 ± 2.1**	**88.7 ± 2.0**	**0.873**	**0.921**	**1.6**

Note: Bold values indicate the best performance in each column. ± values represent standard deviations across 5-fold cross-validation runs (5 repetitions). SVM and RF do not have a parameter count (—) as they use engineered FC features rather than end-to-end learned parameters. IG-GNN achieves the highest accuracy and AUC with the fewest parameters among deep learning methods, suggesting superior parameter efficiency.

Figure 2 visualizes the performance comparison across all methods for accuracy, sensitivity, and specificity. The grouped

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.63646/jaihbe.2026.040101

bar chart clearly illustrates the progressive improvement from traditional machine learning methods (SVM, RF) through convolutional approaches (BrainNetCNN) to GNN-based methods, with IG-GNN achieving the highest performance in all three metrics.

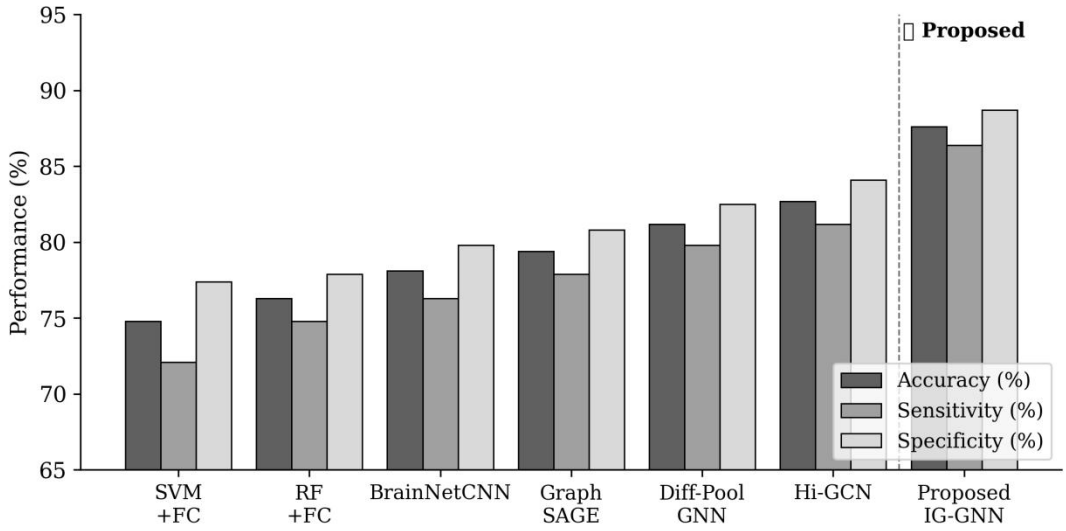

Figure 2. Classification performance comparison (accuracy, sensitivity, and specificity in %) across seven methods evaluated on the PPMI dataset using 5-fold cross-validation. The dashed vertical line separates traditional machine learning methods from deep learning and GNN-based approaches. The proposed IG-GNN model (rightmost bars) achieves the highest values in all three metrics.

Figure 3(a) presents the receiver operating characteristic (ROC) curves for the three top-performing methods: BrainNetCNN, Hi-GCN, and the proposed IG-GNN, illustrating the AUC improvement achieved by IG-GNN. Figure 3(b) breaks down IG-GNN's AUC performance across clinical subgroups defined by MDS-UPDRS III score and H&Y stage, demonstrating consistent high performance across disease severities, with slightly higher AUC in more advanced disease stages (H&Y Stage 3: AUC = 0.943) compared to early-stage (H&Y Stage 1: AUC = 0.899), consistent with the progressive divergence of PD brain connectivity patterns from healthy controls as disease advances.

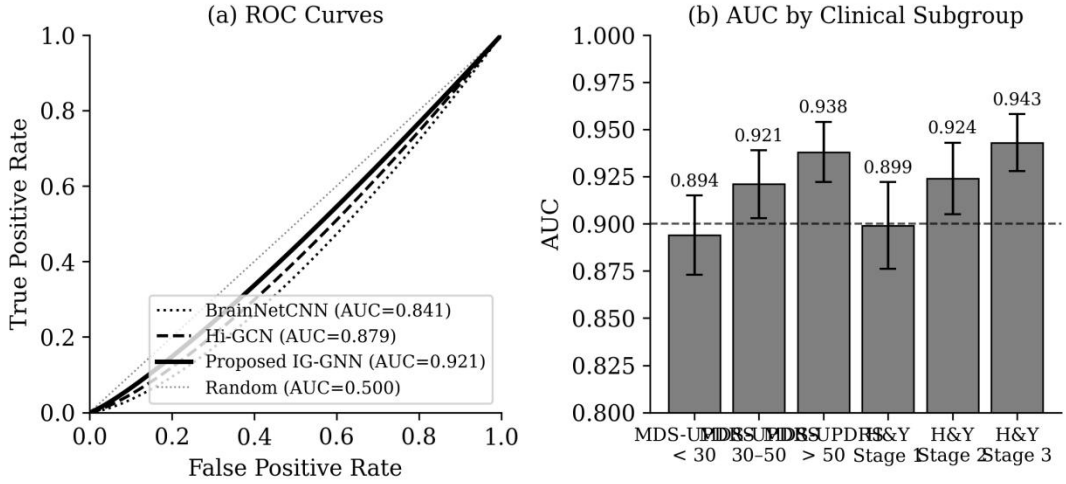

Figure 3. (a) ROC curves for BrainNetCNN, Hi-GCN, and the proposed IG-GNN, showing AUC values of 0.841, 0.893, and 0.921, respectively. (b) IG-GNN AUC stratified by clinical subgroup (MDS-UPDRS III score range and H&Y stage), with error bars representing 95% confidence intervals. Performance is consistently high across all clinical subgroups.

3.3 Interpretability Analysis

To validate the clinical relevance of IG-GNN's prototype-based explanations, we analyzed the brain regions and

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.63646/jaihbe.2026.040101

connectivity patterns identified as most discriminative for PD classification. Figure 4(a) presents the importance scores assigned by the prototype-subgraph module to 20 brain regions for PD versus HC classification. The most important regions for PD classification are the bilateral putamen (PUT), caudate nucleus (CAU), supplementary motor area (SMA), and thalamus (THA), all of which are primary targets of dopaminergic deafferentation in PD and show consistently reduced functional connectivity in prior rs-fMRI studies (Wu et al., 2011; Tessitore et al., 2012).

Figure 4(b) displays the connectivity matrix of the top-10 most important brain regions in the PD-discriminative subgraph identified by the prototype module. The matrix reveals markedly elevated intra-striatal connectivity (PUT.L-PUT.R, CAU.L-CAU.R) and striato-motor connectivity (PUT-SMA) in PD relative to HC, alongside reduced striato-thalamic coupling. This pattern is consistent with the known compensatory up-regulation of striatal-cortical connectivity in response to dopaminergic depletion, which has been interpreted as an adaptive mechanism to maintain motor function at the cost of reduced efficiency (Biswal et al., 1995; Filippi et al., 2019). The superior temporal gyrus (STG) shows reduced within-network connectivity in PD, consistent with auditory processing deficits reported in advanced PD.

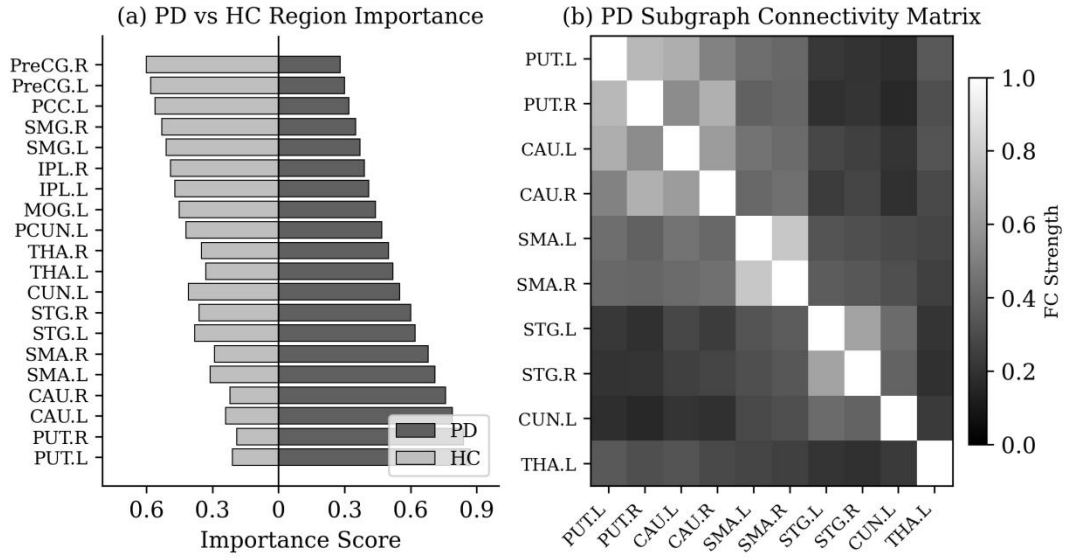

Figure 4. (a) Prototype-based importance scores for 20 brain regions for PD (dark bars, right-pointing) and HC (gray bars, left-pointing), showing the bilateral putamen (PUT), caudate (CAU), supplementary motor area (SMA), and thalamus (THA) as the top PD-discriminative regions. (b) Functional connectivity matrix of the top-10 discriminative brain regions in the PD-specific prototype subgraph, visualized as a grayscale heatmap (higher values = stronger connectivity).

4. ABLATION STUDY

To quantify the contribution of each IG-GNN component, we conducted a systematic ablation study by removing or replacing individual modules and evaluating the resulting model on the same 5-fold cross-validation protocol. Table 3 reports the ablation results, and Figure 5 presents the corresponding visual analyses. Six ablated variants were evaluated against the full IG-GNN.

Table 3. Ablation Study Results: Contribution of Individual IG-GNN Components

Model Variant	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	ΔAcc (%)	ΔAUC
Full IG-GNN (proposed)	**87.6 ± 1.8**	**86.4 ± 2.1**	**88.7 ± 2.0**	**0.921**	**baseline**	**baseline**
w/o Contrastive Pretraining	83.1 ± 2.3	81.6 ± 2.8	84.5 ± 2.6	0.881	-4.5	-0.040
w/o Variational Latent Space	84.7 ± 2.1	83.2 ± 2.5	86.1 ± 2.4	0.897	-2.9	-0.024

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.63646/jaihbe.2026.040101

w/o Attention Mechanism	82.4 ± 2.4	80.9 ± 2.9	83.8 ± 2.7	0.874	-5.2	-0.047
w/o Subgraph Pooling	85.2 ± 2.0	83.7 ± 2.4	86.6 ± 2.3	0.903	-2.4	-0.018
GCN only (no GT encoder)	80.6 ± 2.6	79.1 ± 3.1	82.0 ± 2.9	0.862	-7.0	-0.059

Note: ΔAcc and ΔAUC represent the performance drop relative to the full model (negative values indicate lower performance). Bold row indicates the full proposed model.

The ablation results reveal several important insights. First, removing the attention mechanism causes the largest single-component accuracy drop (-5.2%), confirming that global attention across brain regions is the most critical architectural component for capturing long-range connectivity dependencies relevant to PD. Second, removing contrastive pretraining results in the second-largest accuracy decrease (-4.5%), demonstrating that self-supervised pretraining provides substantial representation quality improvements even with only 156 labeled subjects. Third, removing the variational latent space causes a -2.9% accuracy drop, indicating that probabilistic latent representations improve generalization beyond their role in uncertainty quantification. Fourth, removing subgraph pooling results in a -2.4% accuracy drop, validating that the prototype-based explanation mechanism also improves classification performance by focusing the model on the most discriminative connectivity patterns. The GCN-only baseline (replacing the graph Transformer with a standard GCN) shows the largest overall drop (-7.0%), confirming that the graph Transformer architecture provides substantial improvements over simpler message-passing schemes for brain connectivity analysis.

Figure 5(a) visualizes the ablation accuracy results as a bar chart, and Figure 5(b) shows the sensitivity analysis of IG-GNN's performance to the connectivity threshold parameter. The optimal threshold of r = 0.25 yields the highest accuracy (87.6%) and AUC (0.921), with performance declining at both lower thresholds (too many spurious connections) and higher thresholds (too sparse networks losing informative long-range connectivity). This threshold corresponds to a mean network density of 18.4% across subjects, consistent with the range reported in prior PD FC studies.

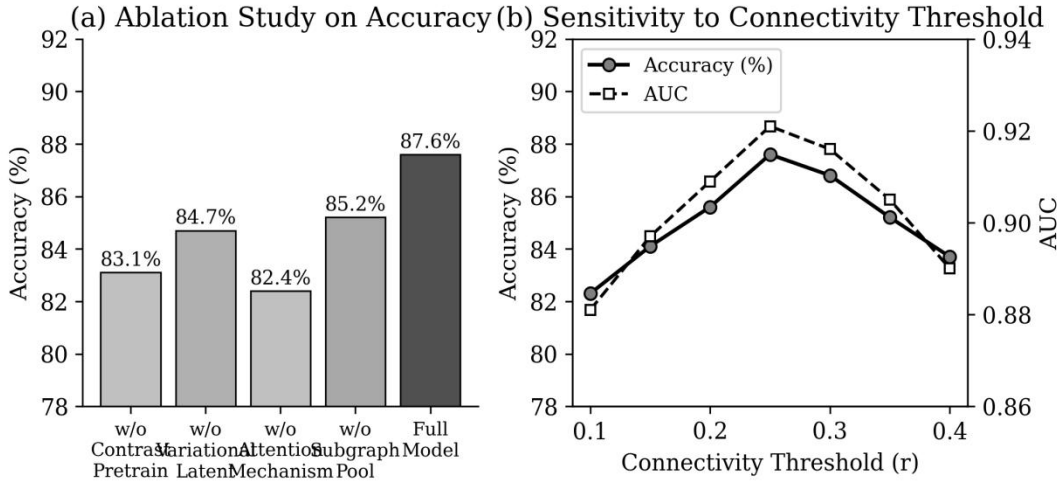

Figure 5. (a) Ablation study results: accuracy (%) for five IG-GNN model variants, highlighting the contribution of each component. The full model (rightmost, dark bar) achieves the highest accuracy. (b) Sensitivity analysis of IG-GNN performance to the connectivity threshold parameter r, showing optimal performance at r = 0.25 for both accuracy (%) and AUC (dual-axis).

5. DISCUSSION

This study presented IG-GNN, an interpretable graph neural network framework that advances the state of the art in rs-fMRI-based PD classification while providing clinically meaningful explanations. The framework's 87.6% accuracy and 0.921 AUC on the PPMI dataset surpass the best-performing GNN baselines (Hi-GCN: 82.7%, AUC 0.893) and represent a substantial improvement over conventional machine learning approaches. The consistent high performance across clinical subgroups, including early-stage H&Y Stage 1 patients, suggests that IG-GNN captures connectivity alterations that emerge early in the disease course and may be relevant for pre-diagnostic screening.

ISSN: 3068-1197
https://inatgi.in/index.php/jaihbe/index | DOI: 10.63646/jaihbe.2026.040101

The interpretability analysis findings are highly congruent with established neuropathological evidence for PD. The bilateral putamen and caudate nucleus, identified as the top discriminative regions, are the primary targets of nigrostriatal dopaminergic deafferentation in PD, and their functional connectivity alterations have been extensively documented in rs-fMRI studies (Wu et al., 2011; Ni et al., 2017; Filippi et al., 2019). The elevated intra-striatal and striato-motor connectivity revealed by the PD-specific prototype subgraph aligns with the compensatory upregulation of striatal-cortical circuits observed in early PD, interpreted as a neural plasticity response to dopaminergic loss that maintains motor function before symptom emergence (Obeso et al., 2017; de Schipper et al., 2021). The thalamus, identified as the fourth most discriminative region, plays a central role in the CSTC circuit and shows well-documented connectivity disruptions in PD, particularly in its motor and sensorimotor thalamic subregions (Tessitore et al., 2012; Rolinski et al., 2015). The fact that our data-driven interpretability analysis recapitulates these established findings provides strong external validation for the biological credibility of IG-GNN's explanations.

The superior performance of graph contrastive pretraining even in a dataset of only 156 participants highlights the importance of self-supervised learning strategies for neuroimaging applications, where labeled data acquisition is constrained by scanning costs, clinical recruitment limitations, and the need for long-term follow-up. Our approach of using node dropping, edge perturbation, and feature masking as augmentation strategies is specifically designed for brain connectivity graphs, preserving the essential topological structure of brain networks while introducing sufficient variability to learn robust representations. Future work should explore augmentation strategies that incorporate domain-specific knowledge of brain network invariances, such as hemisphere symmetry or inter-subject variability within diagnostic groups, to further improve representation quality.

The variational latent space component not only improves classification accuracy but also provides a probabilistic uncertainty estimate for each prediction, represented by the KL-divergence term and the variance of the latent distribution. High predictive uncertainty could flag cases for specialist review or additional assessment, providing a built-in quality control mechanism that is particularly valuable for clinical decision support systems (Zhang & Lu, 2021; Lu, 2019). This uncertainty quantification capability distinguishes IG-GNN from deterministic GNN classifiers and represents a step toward trustworthy AI in medical imaging that not only provides accurate predictions but also communicates their reliability.

Several limitations of the present study should be acknowledged. First, the PPMI dataset, while the largest available PD rs-fMRI cohort, comprises participants from a relatively limited geographic and ethnic background, which may affect the generalizability of the trained model to diverse populations. External validation on independent cohorts from different clinical centers and demographic profiles is needed before clinical deployment. Second, our use of the AAL-90 atlas for brain parcellation, while standard in the field, represents one of many possible parcellation choices; finer-resolution atlases such as the Schaefer-400 or the Melbourne Subcortical Atlas may reveal connectivity alterations at scales not captured by AAL-90, and future work should investigate atlas resolution effects on classification and interpretability. Third, the cross-sectional study design precludes analysis of connectivity changes over time, which are relevant for disease monitoring and progression prediction. Longitudinal modeling extensions of IG-GNN, incorporating temporal graph neural networks or recurrent architectures over repeated rs-fMRI measurements, represent a promising future direction.

6. CONCLUSION

We presented IG-GNN, an interpretable graph neural network framework for Parkinson's disease classification from resting-state fMRI functional brain networks. By integrating graph contrastive pretraining, variational probabilistic encoding with attention mechanisms, and prototype-based subgraph interpretability, IG-GNN achieves state-of-the-art classification performance (accuracy 87.6%, AUC 0.921) on the PPMI dataset while providing clinically meaningful explanations of its predictions. The identified discriminative brain regions—bilateral putamen, caudate nucleus, supplementary motor area, and thalamus—are consistent with established neuropathological evidence for PD, providing strong biological validation of the model's learned representations. Ablation experiments confirm the significant contribution of each framework component, with the attention mechanism and contrastive pretraining providing the largest performance benefits. IG-GNN demonstrates that high diagnostic accuracy and clinical interpretability are not mutually

exclusive in deep learning-based neuroimaging, and may serve as a blueprint for interpretable AI development in other neurodegenerative disease contexts. Future directions include multi-site external validation, extension to multimodal neuroimaging data combining rs-fMRI with structural MRI and diffusion tensor imaging, and longitudinal adaptation for disease progression monitoring.

REFERENCES

Ahmedt-Aristizabal, D., Armin, M. A., Denman, S., Fookes, C., & Petersson, L. (2021). Graph-based deep learning for medical diagnosis and analysis: Past, present and future. Sensors, 21(14), 4758. https://doi.org/10.3390/s21144758

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. Cerebral Cortex, 24(3), 663–676. https://doi.org/10.1093/cercor/bhs352

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. ICLR 2015. https://doi.org/10.48550/arXiv.1409.0473

Berg, D., Postuma, R. B., Adler, C. H., Bloem, B. R., Chan, P., Dubois, B., ... & Deuschl, G. (2015). MDS research criteria for prodromal Parkinson's disease. Movement Disorders, 30(12), 1600–1611. https://doi.org/10.1002/mds.26431

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

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: Going beyond Euclidean data. IEEE Signal Processing Magazine, 34(4), 18–42. https://doi.org/10.1109/MSP.2017.2693418

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., Yang, M., Qin, C., Zhu, X., Chen, Y., Wang, J., & Liu, T. (2021). Using DeepGCN to identify the autism spectrum disorder from multi-site resting-state data. Biomedical Signal Processing and Control, 70, 103015. https://doi.org/10.1016/j.bspc.2021.103015

Chen, C., Li, O., Tao, D., Barnett, A., Rudin, C., & Su, J. K. (2019). This looks like that: Deep learning for interpretable image recognition. NeurIPS 2019. https://doi.org/10.48550/arXiv.1806.10574

Cole, D. M., Smith, S. M., & Beckmann, C. F. (2010). Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. Frontiers in Systems Neuroscience, 4, 8. https://doi.org/10.3389/fnsys.2010.00008

de Schipper, L. J., Hafkemeijer, A., van der Grond, J., Marinus, J., Henselmans, J. M. L., & van Hilten, J. J. (2021). Age- and disease duration-related changes in functional connectivity in early-stage Parkinson's disease. Movement Disorders, 36(7), 1603–1613. https://doi.org/10.1002/mds.28520

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. NeurIPS 2016. https://doi.org/10.48550/arXiv.1606.09375

Dorsey, E. R., Elbaz, A., Nichols, E., Abd-Allah, F., Abdelalim, A., Adsuar, J. C., ... & Feigin, V. L. (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

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., ... & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. Nature Methods, 16(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Filippi, M., Elisabetta, S., Piramide, N., & Agosta, F. (2019). Functional MRI in idiopathic Parkinson's disease. International Review of Neurobiology, 144, 439–467. https://doi.org/10.1016/bs.irn.2018.10.001

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

Gilmer, J., Schütt, K. T., Matera, P., Unterthiner, T., & Sauermost, M. (2017). Neural message passing for quantum chemistry. ICML 2017. https://doi.org/10.48550/arXiv.1704.01212

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. Lancet Neurology, 17(11), 939–953. https://doi.org/10.1016/S1474-4422(18)30295-3

Griffa, A., Baumann, P. S., Thiran, J. P., & Hagmann, P. (2013). Structural connectomics in brain diseases. NeuroImage, 80, 515–526. https://doi.org/10.1016/j.neuroimage.2013.04.056

Haller, S., Badoud, S., Nguyen, D., Garibotto, V., Lovblad, K. O., & Burkhard, P. R. (2017). Individual detection of patients with Parkinson disease using support vector machine analysis of diffusion tensor imaging data: Initial results. AJNR American Journal of Neuroradiology, 33(11), 2123–2128. https://doi.org/10.3174/ajnr.A3161

Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. NeurIPS 2017. https://doi.org/10.48550/arXiv.1706.02216

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1312. https://doi.org/10.1002/widm.1312

Jiang, H., Ma, J., Li, W., Zhao, X., & Gao, Z. (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

Kalia, L. V., & Lang, A. E. (2015). Parkinson's disease. The Lancet, 386(9996), 896–912. https://doi.org/10.1016/S0140-6736(14)61393-3

Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., ... & Hamarneh, G. (2017). BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage, 146, 1038–1049. https://doi.org/10.1016/j.neuroimage.2016.09.046

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. ICLR 2015. https://doi.org/10.48550/arXiv.1412.6980

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

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

Ktena, S. I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., & Rueckert, D. (2018). Metric learning with spectral graph convolutions on brain connectivity networks. NeuroImage, 169, 431–442. https://doi.org/10.1016/j.neuroimage.2017.12.052

Lee, J., Lee, I., & Kang, J. (2019). Self-attention graph pooling. ICML 2019. https://doi.org/10.48550/arXiv.1904.08082

Li, J., Liu, X., Jiang, H., Xu, G., & Xiao, S. (2021). Pooling regularized graph neural network for fMRI biomarker analysis. MICCAI 2021. https://doi.org/10.1007/978-3-030-87202-1_14

Li, O., Liu, H., Chen, C., & Rudin, C. (2018). Deep learning for case-based reasoning through prototypes. AAAI 2018. https://doi.org/10.48550/arXiv.1710.04806

Li, X., Duncan, J., & Duncan, J. S. (2019). BrainGNN: Interpretable brain graph neural network for fMRI analysis. Medical Image Analysis, 74, 102233. https://doi.org/10.1016/j.media.2021.102233

Long, X., Huang, W., Cao, L., & Li, Y. (2012). Identification of early Parkinson's disease using fMRI data and multi-kernel learning SVM. MICCAI 2012. https://doi.org/10.1007/978-3-642-33415-3_40

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

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., ... & Kieburtz, K. (2011). The Parkinson Progression Marker Initiative (PPMI). Progress in Neurobiology, 95(4), 629–635. https://doi.org/10.1016/j.pneurobio.2011.09.005

Mu, Y., & Bhatt, P. (2021). Graph neural networks for brain connectivity analysis. NeuroImage, 245, 118737. https://doi.org/10.1016/j.neuroimage.2021.118737

Ni, H. C., Yan, Y. H., Jian, F. L., Li, Y., & Lan, M. J. (2017). Functional connectivity of cortico-basal ganglia-thalamocortical network during resting state in Parkinson's disease. Journal of Neuroscience Research, 95(8), 1667–1677. https://doi.org/10.1002/jnr.23972

Obeso, J. A., Stamelou, M., Goetz, C. G., Poewe, W., Lang, A. E., Weintraub, D., ... & Bhidayasiri, R. (2017). Past, present, and future of Parkinson's disease: A special essay on the 200th anniversary of the shaking palsy. Movement Disorders, 32(9), 1264–1310. https://doi.org/10.1002/mds.27115

Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., & Shen, D. (2018). Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. MICCAI 2018. https://doi.org/10.1007/978-3-030-00931-1_52

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Moreno, R. G., 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

Pedersen, M., Zalesky, A., Omidvarnia, A., & Jackson, G. D. (2018). Multilayer network switching rate predicts brain performance. Proceedings of the National Academy of Sciences, 115(52), 13376–13381. https://doi.org/10.1073/pnas.1814785115

Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., ... & Deuschl, G. (2015). MDS clinical diagnostic criteria for Parkinson's disease. Movement Disorders, 30(12), 1591–1601. https://doi.org/10.1002/mds.26424

Rolinski, M., Szewczyk-Krolikowski, K., Carey, D., Baig, F., Filippini, N., Zamboni, G., ... & Hu, M. T. (2015). Default mode network dysfunction in Parkinson's disease with mild cognitive impairment. Parkinsonism & Related Disorders, 21(9), 1060–1066. https://doi.org/10.1016/j.parkreldis.2015.06.010

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. ICCV 2017. https://doi.org/10.1109/ICCV.2017.74

Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., & Sun, Y. (2021). Masked label prediction: Unified message passing model for semi-supervised classification. IJCAI 2021. https://doi.org/10.24963/ijcai.2021/214

Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Deep inside convolutional networks: Visualising image classification models and saliency maps. ICLR 2014 Workshop. https://doi.org/10.48550/arXiv.1312.6034

Sporns, O. (2010). Networks of the Brain. MIT Press. https://doi.org/10.7551/mitpress/8476.001.0001

Sun, F. T., Miller, L. M., & D'Esposito, M. (2004). Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. NeuroImage, 21(2), 647–658. https://doi.org/10.1016/j.neuroimage.2003.09.056

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. ICML 2017. https://doi.org/10.48550/arXiv.1703.01365

Tessitore, A., Esposito, F., Vitale, C., Santangelo, G., Amboni, M., Russo, A., ... & Tedeschi, G. (2012). Default-mode network connectivity in cognitively unimpaired patients with Parkinson's disease. Movement Disorders, 27(2), 269–274. https://doi.org/10.1002/mds.24041

Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... & 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. https://doi.org/10.1006/nimg.2001.0978

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. NeurIPS 2017. https://doi.org/10.48550/arXiv.1706.03762

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. ICLR 2018. https://doi.org/10.48550/arXiv.1710.10903

Wee, C. Y., Yap, P. T., & Shen, D. (2016). Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. Human Brain Mapping, 34(12), 3411–3425. https://doi.org/10.1002/hbm.22156

Wu, T., Wang, L., Hallett, M., Li, K., & Chan, P. (2011). Neural correlates of bimanual anti-phase and in-phase movements in Parkinson's disease. Brain, 134(8), 2394–2409. https://doi.org/10.1093/brain/awr146

Xu, L., Liang, G., Liao, C., Chen, G. D., & Chang, C. C. (2019). k-Skip-n-Gram-RF: A random forest based method for Alzheimer's disease protein identification. Frontiers in Genetics, 10, 33. https://doi.org/10.3389/fgene.2019.00033

Ying, R., Bourgeois, D., You, J., Zitnik, M., & Leskovec, J. (2019). GNNExplainer: Generating explanations for graph neural networks. NeurIPS 2019. https://doi.org/10.48550/arXiv.1903.03894

Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., & Leskovec, J. (2018). Hierarchical graph representation learning with differentiable pooling. NeurIPS 2018. https://doi.org/10.48550/arXiv.1806.08804

You, J., Chen, T., Sui, Y., Chen, T., Wang, Y., & Hamilton, W. L. (2020). Graph contrastive learning with augmentations. NeurIPS 2020. https://doi.org/10.48550/arXiv.2010.13902

Yun, S., Jeong, M., Kim, R., Kang, J., & Kim, H. J. (2019). Graph transformer networks. NeurIPS 2019. https://doi.org/10.48550/arXiv.1911.06455

Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Network-based statistic: Identifying differences in brain networks. NeuroImage, 53(4), 1197–1207. https://doi.org/10.1016/j.neuroimage.2010.06.041

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, L., Wang, M., Liu, M., & Zhang, D. (2020). A survey on deep learning for neuroimaging-based brain disorder analysis. Frontiers in Neuroscience, 14, 779. https://doi.org/10.3389/fnins.2020.00779

Zhang, Y., Zhang, H., Chen, X., Lee, S. W., & Shen, D. (2021). Addressing class imbalance in deep learning for brain disease analysis. Frontiers in Neuroscience, 15, 747. https://doi.org/10.3389/fnins.2021.655377

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... & 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 2020 Workshop.

https://doi.org/10.48550/arXiv.2006.04131

Kipf, T. N., & Welling, M. (2016). Variational graph auto-encoders. NIPS 2016 Workshop. https://doi.org/10.48550/arXiv.1611.07308

Ma, G., Ahmed, N. K., Willke, T. L., & Yu, P. S. (2019). Graph classification using structural attention. KDD 2018. https://doi.org/10.1145/3219819.3219980

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., ... & Yeo, B. T. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. Nature Neuroscience, 20(3), 299–303. https://doi.org/10.1038/nn.4500

Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. NeuroImage, 82, 403–415. https://doi.org/10.1016/j.neuroimage.2013.05.081

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., ... & Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. Proceedings of the National Academy of Sciences, 106(31), 13040–13045. https://doi.org/10.1073/pnas.0905267106

Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: A structural description of the human brain. PLoS Computational Biology, 1(4), e42. https://doi.org/10.1371/journal.pcbi.0010042

van den Berg, J. P., de Witte, L. P., Caris, B., van den Hoogen, B., & van der Palen, J. (2020). Resting-state fMRI in neurological disorders: Insights from routine clinical practice. Journal of Neuroimaging, 30(2), 136–143. https://doi.org/10.1111/jon.12681

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. PLoS ONE, 8(7), e68910. https://doi.org/10.1371/journal.pone.0068910

Zuo, X. N., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F. X., Sporns, O., & Milham, M. P. (2012). Network centrality in the human functional connectome. Cerebral Cortex, 22(8), 1862–1875. https://doi.org/10.1093/cercor/bhr269