SG-VIB: Salience-Gated Variational Information Bottleneck for Robust Invariant Brain Graph Learning

Zhaolong Yu[1], Chenghao Zhang[1], Li Xue[1], Zhixin Lin[1], Yebin Huang[1], Shengtao Xiao[1], Lanyan Xue[1,2], Riqing Chen[2,3], and Qiufeng Chen[1,2⋆]

[1] College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China
{12411093003,52411049017,52311049004}@fafu.edu.cn; lzx0717d@163.com
{huangyebin,52411049006,000Q101071}@fafu.edu.cn; chenqiufeng0204@126.com
[2] Digital Fujian Research Institute of Big Data for Agriculture and Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China
riqing.chen@fafu.edu.cn
[3] Fujian Key Lab of Agricultural IOT Applications, Sanming University, Sanming 365004, Fujian, China

Abstract. Graph Neural Networks (GNNs) have shown promise in brain disorder diagnosis but remain vulnerable to noise and multi-site heterogeneity in functional connectivity data. Existing invariant graph learning methods extract stable substructures yet largely overlook sample-dependent representation uncertainty. We propose SG-VIB, a probabilistic framework built on a causal-inspired edge disentanglement backbone. Its "Filter-then-Compress" mechanism uses Salience Gating to modulate disease-relevant graph features before a Variational Information Bottleneck maps them into a stochastic latent space. Uncertainty-Aware Reliable Contrastive Learning (U-RCL) then converts posterior dispersion into detached reliability weights during optimization. Experiments on the multi-site ABIDE I cohort and an ADNI subset show consistent improvements over representative baselines.

Keywords: Brain Network Analysis · Invariant Graph Learning · Variational Inference · Uncertainty Quantification

1 Introduction

Graph Neural Networks (GNNs) are a powerful paradigm for modeling functional brain connectivity and diagnosing neurological disorders [1, 2]. However, fMRI data is susceptible to high-dimensional noise and multi-site heterogeneity [3], and graph models may rely on spurious correlations that fail to generalize across environments [4]. Recent invariant graph learning approaches therefore seek substructures that remain stable across environments [5, 21].

⋆ Corresponding author.

Despite their promise, these methods typically use deterministic embeddings [5, 21] and do not represent sample-dependent uncertainty. This limitation is consequential for noisy, heterogeneous neuroimaging data, where a single point embedding cannot express how concentrated or dispersed a sample representation is. The Information Bottleneck (IB) offers a principled framework for filtering nuisance information [6], while its variational form enables stochastic latent representations [7]. However, indiscriminate compression may remove subtle disease-related patterns together with noise.

Invariant subgraph extraction identifies cohort-level stable patterns, whereas uncertainty modeling characterizes how concentrated each sample's latent representation is. A deterministic invariant encoder may remain overconfident on atypical observations, while unselective stochastic compression may weaken sparse diagnostic patterns. This motivates coupling salience modulation with probabilistic encoding.

Accordingly, we propose SG-VIB. Built on a causal-inspired edge disentanglement backbone, it applies Salience Gating before VIB to obtain stochastic latent representations without indiscriminately compressing graph features. U-RCL then converts posterior variance into detached reliability weights for contrastive learning. The resulting contributions are a probabilistic extension of invariant brain graph learning, a Filter-then-Compress design, and uncertainty-aware contrastive optimization.

2 Methodology

As illustrated in Fig. 1, SG-VIB combines a deterministic edge-disentanglement backbone with Salience Gating, VIB, U-RCL, and consistency regularization. The learned high-score subgraphs are treated as putatively invariant, disease-relevant associations rather than formally identified causal subgraphs.

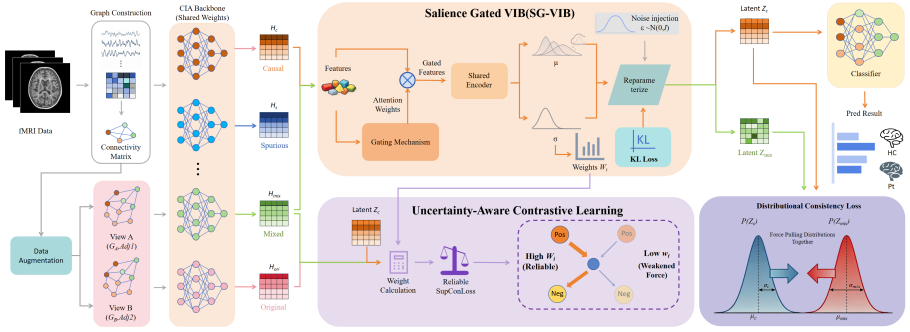


Fig. 1. SG-VIB overview. A deterministic backbone refines the adjacency, disentangles putatively invariant and complementary subgraphs, and encodes $\mathbf{h}_g$. Salience Gating and VIB produce a stochastic representation $(\boldsymbol{\mu}, \boldsymbol{\sigma})$, whose posterior dispersion weights U-RCL.

2.1 Deterministic Backbone

A lightweight BrainNetCNN-style front end [32], comprising Edge-to-Edge (E2E) and Edge-to-Node (E2N) convolutions, first extracts topology-aware features from the functional connectivity matrix. Following [21], learned edge attention then separates high-score, putatively invariant edges from complementary edges. A GCN [15] with attention pooling encodes the former into $\mathbf{h}_g \in \mathbb{R}^d$.

These stages serve distinct purposes: E2E aggregates row- and column-wise context among incident connections while preserving the matrix structure, E2N summarizes the resulting edge responses into node-level features for graph construction, and edge attention identifies high-score connections within the constructed topology. Consequently, $\mathbf{h}_g$ summarizes the selected graph pathway before probabilistic processing. The following modules therefore model uncertainty in this graph-level representation rather than uncertainty in individual raw correlations.

2.2 Salience Gating Mechanism

To modulate residual noise before stochastic compression, a two-layer MLP generates a saliency mask $\mathbf{m} \in [0,1]^d$ with $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{d\times d}$:

$$\mathbf{m} = \sigma\left(\mathbf{W}_2 \cdot \mathrm{ReLU}\left(\mathrm{LN}(\mathbf{W}_1\mathbf{h}_g + \mathbf{b}_1)\right) + \mathbf{b}_2\right), \tag{1}$$

where $\mathrm{LN}(\cdot)$ is Layer Normalization [8] and $\tilde{\mathbf{h}} = \mathbf{h}_g \odot \mathbf{m}$. We initialize $\mathbf{b}_2 = 2.0$ so that the gate is initially open.

The graph-specific mask modulates representation dimensions rather than thresholding edges. Its sigmoid range preserves gradient flow, while the positive initialization makes the initial transformation close to identity and allows gradual feature attenuation during training.

2.3 Variational Information Bottleneck: Stochastic Embedding

Following IB [6], the gated feature is encoded into stochastic latent parameters:

$$\mathbf{h}_e = \mathrm{Encoder}(\tilde{\mathbf{h}}), \quad \boldsymbol{\mu} = \mathbf{h}_g + \mathrm{MLP}_\mu(\mathbf{h}_e), \quad \log\boldsymbol{\sigma}^2 = \mathrm{MLP}_\sigma(\mathbf{h}_e). \tag{2}$$

The residual path anchors $\boldsymbol{\mu}$ to $\mathbf{h}_g$ for optimization stability, while its learned offset and variance remain conditioned on $\tilde{\mathbf{h}}$; hence, the gate modulates but does not block the posterior mean. Using reparameterization [9], $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\epsilon} \odot \boldsymbol{\sigma}$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $q_\phi(\mathbf{Z}|G) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$. The VIB objective is

$$\mathcal{L}_{\mathrm{VIB}} = \mathcal{L}_{\mathrm{CE}}(g(\mathbf{z}), y) + \beta(t) \cdot D_{\mathrm{KL}}\left[q_\phi(\mathbf{Z}|G) \| p(\mathbf{Z})\right], \tag{3}$$

where g is the classifier and $p(\mathbf{Z})$ is a standard Gaussian. To avoid posterior collapse, $\beta(t)$ increases linearly from 10^{-4} to 10^{-2} over the first 30 epochs.

Here, $\boldsymbol{\mu}$ carries the central discriminative representation, whereas $\boldsymbol{\sigma}^2$ describes the spread of plausible latent realizations. Both depend on the salience-conditioned encoder, while the residual affects only the mean. Cross-entropy preserves label information and KL divergence regularizes the posterior.

2.4 Uncertainty-Aware Reliable Contrastive Learning (U-RCL)

For graph G_i, posterior variance controls the dispersion of latent draws around $\boldsymbol{\mu}_i$. We therefore define sample-level uncertainty as mean log-variance:

$$u_i = \frac{1}{d}\sum_{j=1}^{d} \log \sigma_{ij}^2. \tag{4}$$

The log scale matches the encoder output and averaging normalizes over dimensions. This score measures latent posterior dispersion, from which we compute

$$w_i = \frac{\exp(-u_i)}{\mathbb{E}[\exp(-u)] + \delta}, \quad \text{with } w_i \leftarrow \text{stop\_grad}(w_i), \tag{5}$$

where the expectation is the mini-batch mean and $\delta > 0$ ensures numerical stability. Larger dispersion yields a smaller anchor weight. Detachment prevents the model from reducing a difficult sample's contribution merely by inflating its variance.

Mini-batch normalization keeps the average weight close to one, so w_i changes the relative emphasis among anchors rather than uniformly shrinking the loss. High-dispersion samples still contribute to classification and VIB and may occur in other anchors' positive or negative sets. Thus, U-RCL performs soft reliability-aware reweighting rather than hard sample selection.

The uncertainty-weighted supervised contrastive loss [10] is defined as:

$$\mathcal{L}_{\text{U-RCL}} = \sum_{i\in I} \frac{-1}{|P(i)|} \sum_{p\in P(i)} w_i \cdot \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p/\tau)}{\sum_{a\in A(i)} \exp(\mathbf{z}_i \cdot \mathbf{z}_a/\tau)}, \tag{6}$$

where $\mathbf{z}$ is ℓ_2-normalized, I indexes the mini-batch, $P(i) = \{p \neq i : y_p = y_i\}$, and $A(i) = I \setminus \{i\}$. Anchors without positives are omitted.

2.5 Optimization

For consistency regularization, we draw two conditionally independent views from the same posterior:

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \qquad \mathbf{z}' = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}', \quad \boldsymbol{\epsilon}, \boldsymbol{\epsilon}' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{7}$$

Thus, $\mathbf{z}' - \mathbf{z} \sim \mathcal{N}(\mathbf{0}, 2\,\text{diag}(\boldsymbol{\sigma}^2))$ and its posterior-adaptive magnitude satisfies $\mathbb{E}[\|\mathbf{z}' - \mathbf{z}\|_2^2 \mid G] = 2\sum_j \sigma_j^2$. We minimize

$$\mathcal{L}_{\text{cons}} = \|\mathbf{z} - \mathbf{z}'\|_2^2. \tag{8}$$

The final objective is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{VIB}} + \lambda_1 \mathcal{L}_{\text{U-RCL}} + \lambda_2 \mathcal{L}_{\text{cons}}, \tag{9}$$

where validation-fold grid search selects $\lambda_1 \in \{0.1, 0.5, 1.0\}$ and $\lambda_2 \in \{0.01, 0.1, 0.5\}$.

Together, VIB learns a predictive compressed posterior, U-RCL organizes class-consistent neighborhoods with dispersion-aware weights, and consistency regularization limits variation between posterior draws. Since $\boldsymbol{\sigma}$ determines the perturbation scale, this constraint adapts to each sample without another noise hyperparameter.

3 Experiments and Results

3.1 Experimental Setup

Datasets and Preprocessing. We used two benchmark datasets: (1) ABIDE I [1], with 1,035 subjects (505 ASD, 530 TD) from 20 acquisition cohorts contributed by 17 institutions after quality control; and (2) an ADNI [11, 27] rs-fMRI subset selected for this study, with 92 subjects (31 AD, 61 NC) scanned at 3.0T. Data were processed using DPARSF [12] (slice timing, realignment, normalization, and band-pass filtering). Nodes were defined by the AAL atlas [13]; Pearson correlations formed the functional connectivity matrix, from which a sparse k-NN graph ($k = 5$) was constructed.

Baselines and Implementation Details. We compared SG-VIB with RF [14]; GCN [15], GAT [16], GIN [17], and GRAND [23]; BrainGB [18], Contrasformer [24], and ℂℂWBrainNet [25]; and the causal/invariant methods DIR-GNN [5], CI-GNN [20], BrainIB [19], CIA-GCL [21], and DBGSL [26].

Models were implemented in PyTorch [29] and trained for 150 epochs using Adam [30], an initial learning rate of 5×10^{-4}, and cosine annealing [31]. We used $d = 128$, batch size 32, two GCN layers with attention readout, and $\tau = 0.5$.

We used diagnosis-stratified, subject-level 10-fold cross-validation, with each subject tested once (103–104 ABIDE and 9–10 ADNI subjects per fold). Diagnosis stratification maintains comparable class proportions, and subject-level separation prevents participant overlap between training and testing. Site was not a stratification variable, so its fold-wise distribution follows the pooled cohort rather than being fixed or balanced. All methods shared preprocessing, graph construction, and subject splits; hyperparameters were selected on validation data separated from each test fold. Thus, the reported mean and standard deviation measure subject-held-out variation under a common pooled-cohort protocol rather than leave-one-site-out generalization.

3.2 Results and Analysis

Table 1 shows consistent gains across threshold-dependent and ranking metrics. On ABIDE I, SG-VIB reaches 74.8% ACC, 75.6% AUC, and 76.3% F1, improving ACC and AUC over CIA-GCL by 2.2 points. On ADNI, it obtains 78.1% AUC versus 71.7% for the second-best method, together with the highest ACC and F1, although the smaller cohort ($N = 92$) yields larger fold-wise variance. Its higher Recall is accompanied by lower Precision than CIA-GCL on ABIDE I (73.2% vs.

74.5%) and ADNI (63.3% vs. 83.2%), indicating a sensitivity–precision trade-off relevant to screening settings.

Table 1. Classification performance comparison (Mean ± Std %). Best results are in **bold**, and the second-best results are underlined.

Method	ACC	AUC	F1	Recall	Precision
Dataset 1: ABIDE I					
RF	62.1±4.8	65.8±6.2	63.3±3.7	63.2±3.6	63.3±5.3
GCN	66.1±6.2	66.2±6.2	65.2±8.0	65.5±13.2	67.6±8.2
GAT	68.7±3.9	68.5±3.5	69.4±4.2	72.2±8.3	64.5±4.2
GIN	67.3±4.2	67.8±4.1	67.3±4.2	68.9±5.2	66.3±6.9
DIR-GNN	68.7±1.6	67.3±1.3	63.8±2.5	57.6±3.2	73.3±10.2
BrainGB	70.5±4.2	74.3±5.0	70.2±4.2	71.3±6.3	72.6±7.2
Contrasformer	68.9±2.3	70.6±2.6	68.7±2.7	70.9±6.0	67.7±3.9
CI-GNN	71.2±2.2	73.1±3.4	70.2±3.2	72.3±4.3	70.1±2.8
GRAND	71.1±4.3	72.8±4.2	71.4±6.9	72.4±12.4	71.9±4.6
ℂℂWBrainNet	67.2±2.0	72.3±2.1	66.3±3.7	69.2±5.5	63.1±5.8
BrainIB	70.1±2.4	72.8±3.2	70.4±2.6	70.1±4.1	71.2±3.6
CIA-GCL	72.6±3.1	73.4±4.6	72.2±3.0	71.2±7.8	**74.5±6.5**
DBGSL	72.2±4.1	73.2±3.8	73.7±4.5	76.6±9.0	71.8±5.2
SG-VIB (Ours)	**74.8±3.0**	**75.6±4.5**	**76.3±3.6**	**80.5±8.6**	73.2±4.6
Dataset 2: ADNI					
RF	65.2±13.5	54.9±20.1	41.2±31.8	38.3±30.6	45.5±35.1
GCN	70.3±8.7	49.0±17.1	31.0±32.4	30.0±31.4	33.3±35.7
GAT	74.5±12.0	56.2±30.8	35.3±37.5	30.0±31.0	46.6±47.6
GIN	68.0±4.7	42.0±16.6	15.8±21.8	26.6±35.1	18.6±24.0
DIR-GNN	80.7±14.2	71.7±29.3	60.0±44.2	54.0±39.8	56.1±41.1
BrainGB	79.8±8.8	63.6±31.9	46.6±36.4	65.0±43.5	51.4±35.2
Contrasformer	72.3±11.4	53.2±27.8	31.6±34.2	26.5±31.2	40.8±43.1
CI-GNN	79.8±9.8	66.6±24.4	48.3±38.1	63.5±43.4	51.6±36.6
GRAND	77.5±6.2	61.9±31.3	54.6±28.3	53.3±33.1	65.3±36.6
ℂℂWBrainNet	74.5±12.3	54.8±30.0	32.5±36.9	27.3±33.7	42.0±46.5
BrainIB	78.4±10.1	64.8±23.1	50.1±35.9	46.6±37.8	59.8±41.2
CIA-GCL	80.5±7.0	66.6±17.6	64.2±17.4	60.8±29.4	**83.2±18.4**
DBGSL	75.6±10.4	39.6±40.0	40.1±40.3	44.2±45.3	38.5±39.5
SG-VIB (Ours)	**82.0±8.4**	**78.1±10.9**	**68.9±16.0**	**82.5±18.4**	63.3±22.1

3.3 Ablation Study

Table 2 evaluates key components on ABIDE I. The w/o U-RCL variant sets $w_i = 1$ in Eq. 6 while retaining consistency regularization.

The w/o Gate variant already benefits from stochastic encoding and U-RCL, explaining its strong AUC. Adding the gate raises ACC from 74.2% to 74.8% but

Table 2. Ablation study on component contributions (ABIDE I dataset). **Gate**: Salience Gating Mechanism; **VIB**: Variational Information Bottleneck; **U-RCL**: Uncertainty-Aware Reliable Contrastive Learning.

Model Variant	Components			Metrics (%)		
	Gate	VIB	U-RCL	ACC	AUC	F1
Backbone Only				72.6	73.4	72.2
SG-VIB w/o Gate		✓	✓	74.2	77.4	76.8
SG-VIB w/o U-RCL	✓	✓		73.6	75.5	75.9
SG-VIB (Full)	✓	✓	✓	**74.8**	**75.6**	**76.3**

slightly lowers AUC and F1; consistent with the residual path in Eq. 2, the gate modulates rather than exclusively selects information. Replacing U-RCL while retaining VIB and consistency regularization reduces ACC to 73.6%, isolating the benefit of reliability-aware anchor weighting. Overall, the full probabilistic extension improves ACC and AUC over the deterministic backbone by 2.2 points.

3.4 Clinical Interpretability and Discussion

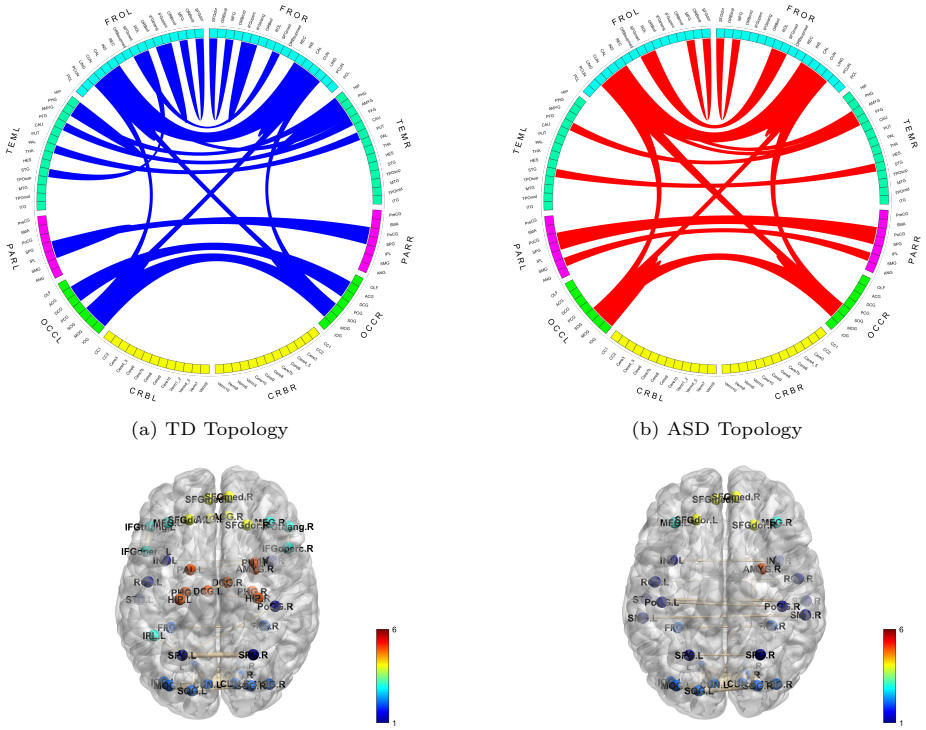
Figure 2 presents the model-selected disease-relevant substructures through chord diagrams and spatial brain maps. These patterns are interpreted as disease-related associations rather than identified causal effects.

Across both views, TD subjects (Fig. 2a, c) exhibit dense, distributed connectivity, whereas ASD subjects (Fig. 2b, d) show sparse long-range links and hyper-connected local clusters, consistent with the "local over-connectivity, global under-connectivity" hypothesis [22].

The two visualization modes are complementary: chord diagrams emphasize group differences in inter-regional topology, while spatial maps locate the highlighted connections on the brain surface. Their qualitative agreement suggests that SG-VIB captures a shift between distributed and locally concentrated connectivity rather than depending on a few isolated regions.

Beyond this cross-view agreement, the visualization clarifies the division of roles within SG-VIB. The edge-disentanglement backbone selects the connections forming the disease-relevant subgraph, whereas Salience Gating and VIB process the resulting graph-level representation. Thus, the visualized topology remains attributable to learned edge scores, while posterior dispersion controls reliability-aware optimization. Read alongside the ablation results, this separation links the quantitative gains to an interpretable reorganization of network connectivity without conflating edge salience with the uncertainty estimate.

From this perspective, examining coordinated network patterns complements the interpretation of individual connections because functional differences may span multiple interacting regions. SG-VIB summarizes these patterns through selected subgraphs, while U-RCL moderates the influence of highly dispersed representations during training. The visualization therefore provides a qualita-



(a) TD Topology (b) ASD Topology

(c) TD Spatial Map (d) ASD Spatial Map

Fig. 2. Model-selected disease-relevant substructures. TD subjects (a, c) show distributed connectivity; ASD subjects (b, d) show local over-connectivity and sparse long-range links. Spatial maps were rendered with BrainNet Viewer [28].

tive counterpart to the classification results and illustrates how reliability-aware learning can preserve coherent group-level connectivity structure.

4 Conclusion

SG-VIB extends invariant brain graph learning through Filter-then-Compress variational encoding and posterior-dispersion-aware contrastive learning, preserving disease-relevant features while recalibrating sample influence. Experiments on ABIDE I and ADNI show consistent gains over deterministic baselines and yield disease-relevant associative patterns.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Grant No. 62301160) and the Natural Science Foundation of Fujian Province (Grant No. 2022J01607).

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

References

1. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol. Psychiatry **19**(6), 659–667 (2014)
2. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., et al.: BrainGNN: Interpretable brain graph neural network for fMRI analysis. Med. Image Anal. **74**, 102233 (2021)
3. Glocker, B., Robinson, R., Castro, D.C., Dou, Q., Konukoglu, E.: Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects. arXiv preprint arXiv:1910.04597 (2019)
4. Li, H., Wang, X., Zhang, Z., Zhu, W.: OOD-GNN: Out-of-distribution generalized graph neural network. IEEE Trans. Knowl. Data Eng. **35**(7), 7328–7340 (2023)
5. Wu, Y.X., Wang, X., Zhang, A., He, X., Chua, T.S.: Discovering invariant rationales for graph neural networks. In: International Conference on Learning Representations (ICLR) (2022)
6. Tishby, N., Pereira, F.C., Bialek, W.: The information bottleneck method. In: Proc. 37th Allerton Conf. Commun. Control Comput., pp. 368–377 (1999)
7. Alemi, A.A., Fischer, I., Dillon, J.V., Murphy, K.: Deep variational information bottleneck. In: International Conference on Learning Representations (ICLR) (2017)
8. Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization. arXiv preprint arXiv:1607.06450 (2016)
9. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: International Conference on Learning Representations (ICLR) (2014)
10. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. In: Larochelle, H., et al. (eds.) NeurIPS 2020. vol. 33, pp. 18661–18673. Curran Associates, Inc. (2020)
11. Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., et al.: The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J. Magn. Reson. Imaging **27**(4), 685–691 (2008)
12. Yan, C.G., Zang, Y.F.: DPARSF: A MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. Front. Syst. Neurosci. **4**, 13 (2010)
13. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage **15**(1), 273–289 (2002)
14. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
15. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017)
16. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (ICLR) (2018)
17. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: International Conference on Learning Representations (ICLR) (2019)
18. Cui, H., Dai, W., Zhu, Y., Kan, X., Gu, A.A.C., Lukemire, J., et al.: BrainGB: A benchmark for brain network analysis with graph neural networks. IEEE Trans. Med. Imaging **42**(2), 493–506 (2023)

19. Zheng, K., Yu, S., Li, B., Jenssen, R., Chen, B.: BrainIB: Interpretable brain network-based psychiatric diagnosis with graph information bottleneck. IEEE Trans. Neural Netw. Learn. Syst. **36**(7), 13066–13079 (2025). https://doi.org/10.1109/TNNLS.2024.3449419
20. Zheng, K., Yu, S., Chen, B.: CI-GNN: A Granger causality-inspired graph neural network for interpretable brain network-based psychiatric diagnosis. Neural Networks **172**, 106147 (2024)
21. Yu, M., Liu, J., Ji, J.: Causal invariance-aware augmentation for brain graph contrastive learning. In: Proceedings of the 42nd International Conference on Machine Learning (ICML). PMLR, vol. 267, pp. 73221–73240 (2025)
22. Courchesne, E., Pierce, K.: Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. Curr. Opin. Neurobiol. **15**(2), 225–230 (2005)
23. Chamberlain, B.P., Rowbottom, J., Gorinova, M.I., Bronstein, M.M., Webb, S., Rossi, E.: GRAND: Graph neural diffusion. In: International Conference on Machine Learning (ICML). PMLR, vol. 139, pp. 1407–1418 (2021)
24. Xu, J., He, K., Lan, M., Bian, Q., Li, W., Li, T., et al.: Contrasformer: A brain network contrastive transformer for neurodegenerative condition identification. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM), pp. 2671–2681 (2024)
25. Jeong, A.Y., Heo, D.W., Suk, H.I.: Complex wavelet-based Transformer for neurodevelopmental disorder diagnosis via direct modeling of real and imaginary components. Med. Image Anal. **109**, 103914 (2026)
26. Campbell, A., Zippo, A.G., Passamonti, L., Toschi, N., Lio, P.: DBGSL: Dynamic brain graph structure learning. In: Proceedings of Medical Imaging with Deep Learning (MIDL). PMLR, vol. 227, pp. 1318–1345 (2024)
27. Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C.R., Jagust, W., et al.: The Alzheimer's Disease Neuroimaging Initiative. Neuroimaging Clin. N. Am. **15**(4), 869–877 (2005). https://doi.org/10.1016/j.nic.2005.09.008
28. Xia, M., Wang, J., He, Y.: BrainNet Viewer: A network visualization tool for human brain connectomics. PLoS ONE **8**(7), e68910 (2013). https://doi.org/10.1371/journal.pone.0068910
29. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al.: PyTorch: An imperative style, high-performance deep learning library. In: NeurIPS. vol. 32, pp. 8024–8035 (2019)
30. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR) (2015)
31. Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. In: International Conference on Learning Representations (ICLR) (2017)
32. Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., et al.: BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage **146**, 1038–1049 (2017). https://doi.org/10.1016/j.neuroimage.2016.09.046