MMBNA: Masked Multiview Brain Network Analysis via Disentangling for Alzheimer's Early Diagnosis with fMRI

Dequan Meng[1], Jie Guo[1], Junze Wang[1], Xiaoming Xi[1], Lishan Qiao[1], Limei Zhang[1,*], and Mingxia Liu[2]

[1] School of Computer Science and Technology, Shandong Jianzhu University, Jinan 250101, Shandong, China
zhanglimei22@sdjzu.edu.cn

[2] Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, NC 27599, USA

Abstract. Alzheimer's disease (AD), as a progressive neurodegenerative disorder, poses a growing global health threat, making early diagnosis imperative. Multiview brain network (BN) analysis from resting-state functional MRI (rs-fMRI) has emerged as a promising approach, where brain regions and their interactions are modeled as nodes and edges across complementary views. However, existing methods have limitations. *First*, they rely on single-measure BNs with fixed nodes and edges, potentially insufficient for capturing complex brain interactions. *Second*, they lack effective separation of view-consistent and view-specific representations, leading to redundancy and reduced generalizability. To address these challenges, we propose a novel Masked Multiview Brain Network Analysis (MMBNA) framework, integrating **multi-measure BNs construction**, **random masking**, and **disentangled representation learning**. Specifically, we first construct multiview BNs via multi-measure connectivity (capturing full/partial/nonlinear correlations) and multi-granularity masking (at node/edge/feature levels), enriching spatio-temporal-topological diversity while preserving semantic similarity. Subsequently, we perform the view-consistent representation learning via cross-view masking, and then a disentangling mechanism is introduced to learn a purer view-specific representation to filter out the redundancy from view-consistent representations, resulting in more compact multiview brain representations. Experiments on the ADNI2 subset of the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, demonstrate the effectiveness of the proposed method, achieving significant improvements in diagnostic accuracy and interpretability compared to state-of-the-art approaches.

Keywords: AD diagnosis · Brain network analysis · Disentangled representation learning · Multiview learning

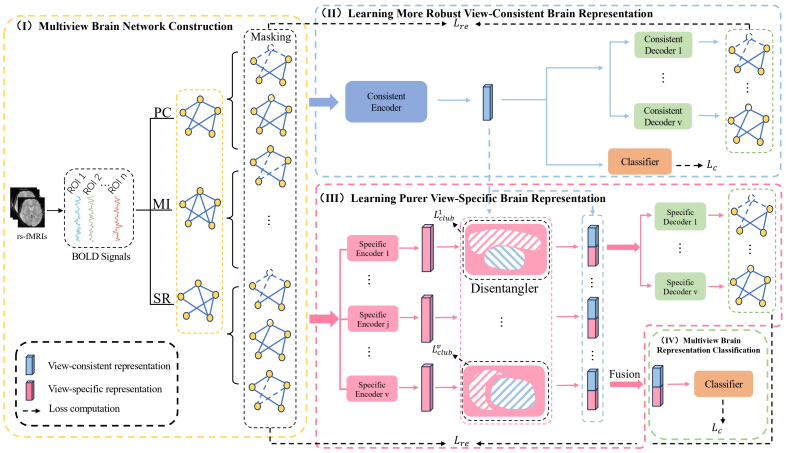

Fig. 1. Illustration of the proposed Masked Multiview Brain Network Analysis (MMBNA) framework, which consists of (I) Multiview Brain Network Construction; (II) Learning More Robust View-Consistent Representation via **Cross-View Masking**; (III) Learning Purer View-Specific Representation through **Disentangling**; (IV) Multiview Brain Representation Classification.

1 Introduction

Alzheimer's disease (AD), a progressive neurodegenerative disorder with complex brain function disturbances, presents major social and healthcare burdens, emphasizing early diagnosis [1, 2]. Resting-state functional magnetic resonance imaging (rs-fMRI) has emerged as a pivotal non-invasive method to understand pathological mechanisms of neurological disorders through the measurement of blood oxygen level-dependent (BOLD) signals [3]. Functional brain network (FBN) analysis, representing BOLD signals in brain regions of interest (ROIs) as nodes and their interactions as edges, has been widely adopted to identify potential biomarkers such as abnormal nodes or edges using machine learning or deep learning approaches, thereby enhancing the accuracy of distinguishing AD patients from healthy controls [4–6].

Recent FBN analysis methods have proven effective in AD diagnosis [6–8]. Most methods rely on single-view brain networks, generally constructed via Pearson correlation (PC), and use deep learning approaches for feature extraction and classification [9, 10]. However, single-view FBN only models one type of correlation, overlooking complex interactions among brain ROIs. To overcome this, multiview FBN methods have been introduced [11, 12]. Some studies generate

multiple views by applying various thresholding strategies on single-measure networks [13, 14], but these methods focus solely on varying FBN sparsity, neglecting the complexity of brain connectivity patterns. Ideally, topologically diverse yet semantically consistent data views are crucial to the performance of multiview FBN analysis. To this end, Zhang et al. [15] and Wang et al. [16] constructed multi-measure brain networks and used linear weighting and tensor decomposition-based views fusion, respectively. However, their fusion methods overlook redundancy between view-consistent and view-specific information, limiting efficiency and performance [17].

To address these challenges, we propose a novel Masked Multiview Brain Network Analysis (MMBNA) framework, which integrates **multi-measure FNB construction**, **masked modeling** with **disentangled representation learning** to derive high-quality brain representations for rs-fMRI-based AD diagnosis. As shown in Fig. 1, multiple FBN views are first constructed using various measures including PC, Sparse Representation (SR) and Mutual Information (MI) to model the full, partial and nonlinear correlations among brain regions, followed by random multi-granularity masking at the node-, feature-, and edge-levels, ensuring the views exhibit both topological diversity and semantic consistency. Subsequently, we learn view-consistent and view-specific representations employing Graph Isomorphism Network (GIN) and Transformer, respectively, while reconstructing masked components in multiviews to capture intrinsic brain representation. Moreover, a disentangling mechanism is introduced to reduce redundancy between view-consistent and view-specific representations, yielding purer and more discriminative representations. Finally, the fused representation is used for AD classification, ensuring both accuracy and interpretability.

2 Methodology

2.1 Masked Multiview Brain Network Construction across Multi-Measure Connectivities

As shown in Stage (I) of Fig. 1, in typical FBN analysis, a brain network is modeled as a graph $G(X, A)$, where $X \in \mathbb{R}^{M \times T}$ represents the node-feature matrix whose rows are average BOLD signals within each ROI, M is the number of ROIs, T is the length of BOLD signals, and $A \in \mathbb{R}^{M \times M}$ is the adjacency matrix denoting relationships among nodes. To capture complex interaction patterns, we employ multiple metrics: PC for full correlations, SR [18] for partial correlations, and MI [19] for nonlinear dependencies to construct brain connectivities. To further enhance view diversity, we introduce different-granularity random masking strategies: **node-based masking** removing subsets of nodes and their edges to alter spatial scales, **edge-based masking** that disturbs topology by masking a proportion of edges, and **feature-based masking** randomly covering a certain ratio of node feature (i.e., original BOLD signals) along the time axis, thereby modifying temporal scales across all the multi-measure brain connectivities. These strategies generate complementary FBN views, ensuring diversity in

spatial, temporal, and topological structures while maintaining semantic similarity. The masking rate, controlled by a hyperparameter, balances structural perturbation and semantic consistency.

Furthermore, it is noteworthy that these masked multiview FBNs would be reconstructed during subsequent brain representation learning, compelling the model to acquire intrinsic discriminative representations, thereby enhancing the generalizability.

2.2 View-Consistent Brain Representation Learning through Masked Modeling

The obtained multiview brain networks with v views, $\left\{G^i(X^i, A^i) \mid i = 1, \ldots, v\right\}$, are fed into a consistent encoder GIN [20] to capture a shared consistent representation z across views, which is subsequently input into consistent decoders for reconstruction, as shown in Stage (II) of Fig. 1. Concretely, for interpretability, we assume the prior of z is a standard normal distribution, $p(z) \sim \mathcal{N}(0, I)$. To facilitate optimization, we apply the reparameterization trick to sample $q_\phi(z \mid G^i) = \mu + \epsilon \cdot \sigma$, where $q_\phi(z \mid G^i)$ is the approximate posterior distribution with the GIN's parameters ϕ, μ and σ are learned parameters, and $\epsilon \sim \mathcal{N}(0, 1)$ is a noise term. Subsequently, z is input into v consistent decoders (MLPs) to predict the adjacency matrix of each view, $\hat{A}^i$. The reconstruction loss across all the masked views is: $L_{\text{re}} = \frac{1}{v}\sum_{i=1}^{v}\left[\text{MSE}(A^i, \hat{A}^i) + (1 - \text{CosSim}(A^i, \hat{A}^i))\right]$, where MSE is Mean Squared Error ensuring precision in local details and CosSim denotes Cosine Similarity preserving consistency in semantic representation.

Moreover, we introduce a classification loss with a MLP classifier: $L_{\text{c}} = -\frac{1}{v}\sum_{i=1}^{v}\left[\alpha(1-\hat{y}^i)^\gamma y^i \log(\hat{y}^i)\right]$, where y^i and $\hat{y}^i$ denote the true label and the predicted probability for the ith view, α is a weighting factor to balance positive and negative, and γ reduces the influence of well-classified samples on the loss. Additionally, a KL divergence term regularizes the consistent representation space to conform to the prior distribution: $L_{KL} = -\frac{1}{v}\sum_{i=1}^{v}\text{KL}(q_\phi(z \mid G^i) \| p(z))$.

The total loss function of Stage (II) combines all the objectives: $L_{II} = \alpha_1 L_{re} + \alpha_2 L_c + L_{KL}$, where α_1 and α_2 are trade-off operators. Notably, the learned parameters of the consistent encoder here will be frozen in the subsequent stage to extract view-consistent representations for disentanglement.

2.3 View-Specific Brain Representation Learning via Disentangling Strategy

As shown in Stage (III) of Fig. 1, to filter out redundancy, we disentangle view-consistent and view-specific representations, obtaining purer view-specific representations. Specifically, for each view, we first extract coarse view-specific representations, denoted as g^i, using a series of Transformer encoders [10] to model long-range dependencies among FBN nodes. Then, considering the intractability of directly estimating mutual information $I(z, g^i)$, we adopt the CLUB estimator [21], which provides a variational upper bound that avoids the explicit computation of marginal distributions. We disentangle g^i from the view-consistent

representation z derived from the frozen consistent encoder GIN, with the following formula:

$$L_{\text{club}} = \frac{1}{v}\sum_{i=1}^{v}(\mathbb{E}_{p(z,g^i)}\left[\log q(g^i \mid z)\right] - \mathbb{E}_{p(z)p(g^i)}\left[\log q(g^i \mid z)\right]) \tag{1}$$

where $p(z, g^i)$ and $p(z)p(g^i)$ represent the joint and marginal distributions of z and g^i, respectively. This estimator minimizes the upper bounds of mutual information, yielding purer view-specific representations $\bar{g}^i$. Subsequently, z and $\bar{g}^i$ are concatenated view-wise and fed into a series of specific decoders (MLPs) for reconstruction and a classifier MLP for prediction, utilizing the same loss functions L_{re}, L_c and L_{KL} in Stage (II). The total loss of Stage (III) combines four objectives:

$$L_{III} = \beta L_{re} + \gamma L_c + \lambda L_{club} + L_{KL} \tag{2}$$

where β, γ and λ are trade-off operators.

2.4 Multiview Brain Representation Classification

During testing, the test data are fed into the trained framework above, where the final fused representation, obtained via averaging, serves for AD diagnosis fully leveraging the complementary information from multiple views, as illustrated in Stage (IV) of Fig. 1.

3 Experiment

Materials and Data Processing. We used the ADNI2 subset of the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset[1] . It comprises 563 subjects distributed as follows: 154 Cognitively Normal (CN), 165 Subjective Memory Complaints (SMC), 145 Mild Cognitive Impairment (MCI), and 99 AD. For performing binary classification, CN (154 samples) was labeled as Class 0, while SMC, MCI, and AD (409 samples in total) were grouped as Class 1. Rs-fMRI data were collected on a 3.0-T Philips scanner (TR = 3000 ms, TE = 30 ms, flip angle = 80°, 48 slices at 3.3 mm thickness, 140 volumes), and the brain was parcellated into 116 regions using the AAL template with 137 time points of BOLD signals.

Experimental Settings. We use 5-fold cross-validation (CV) to split the dataset, training each fold for 60 epochs with a learning rate of 0.001 and a batch size of 32. We use a 3-layer GIN network as the consistent encoder, Transformer with 4-head attention as the view-specific encoder, and 1-layer MLP (hidden dimension = 128, dropout = 0.5) serving as bothe view-consistent and view-specific decoders. Evaluation is based on six metrics: accuracy (ACC), area under the curve (AUC), specificity (SPE), precision (PRE), sensitivity (SEN), and F1-score

[1] `http://adni.loni.usc.edu/`

(F1). All experiments follow a standardized data preprocessing pipeline [22] to ensure fairness and reproducibility.

Competing Methods. We compare the proposed method with three state-of-the-art (SOTA) single-view FBN analysis models (STGCN, BrainGNN, and Transformer) and two multiview methods (MVS-GCN and Zhang et al.'s method). **STGCN** [9]: It employs a three-layer spatio-temporal graph convolutional encoder that integrates 1D temporal convolutions with spatial graph convolutions to model BNs. **BrainGNN** [7]: The model employs ROI-aware graph convolutional layers to capture functional connectivity patterns leveraging the topological and functional information of fMRIs. **Transformer** [10]: This model uses self-attention mechanisms to capture long-range dependencies across brain regions in fMRI data. **MVS-GCN** [13]: It uses multiple FBN views with different sparse structures and integrates graph structure priors to multiview graph embedding learning, resulting in improved classification performance. **Zhang et al.'s Method** [15]: It integrates learning fusion weights for multiple FBN views, constructed by PC, SR, MI, and higher-order measures, with classification using L1-norm support vector machine into a unified framework.

Table 1. Comparison of experimental results on ADNI2 with the best results in bold.

Model	ACC (%)	AUC (%)	SEN (%)	SPE (%)	PRE (%)	F1 (%)
BrainGNN [7]	65.40	61.20	64.85	63.90	66.85	65.83
STGCN [9]	70.11	62.08	86.08	76.85	76.83	80.69
Transformer [10]	75.22	82.70	84.00	57.89	79.75	81.82
MVS-GCN [13]	74.04	62.19	**95.16**	74.24	77.35	85.03
Zhang et al.'s [15]	77.10	84.50	83.90	77.80	83.95	84.58
MMBNA (Ours)	**86.90**	**89.77**	88.71	**81.82**	**93.22**	**90.91**

Experimental Results. As summarized in Table 1, our MMBNA framework achieves state-of-the-art performance on ADNI2, surpassing all competing methods across key metrics. Notably, MMBNA attains the highest accuracy (ACC: 86.90%), outperforming the strongest baseline (Zhang et al. [15]) by 9.80%, and demonstrates significant improvements in AUC (89.77%) and PRE (93.22%). The strong performance of MMBNA can be attributed to two key innovations: (1) By integrating PC (full), SR (partial), and MI (nonlinear) measures with multi-granularity masking (node-, edge-, and feature-level), **MMBNA constructs semantically consistent yet topologically diverse FBN views**, capturing complementary brain connectivity patterns overlooked by other approaches; (2)Through disentangled representation learning, **MMBNA isolates disease-invariant features from measurement specific noise**, reducing redundancy and enhancing discriminative power.

Ablation Study. To evaluate the contribution of each component, we compared MMBNA with its four variants (Table 2): (1) **Only Consistent Learning**, (2)

Table 2. Impact of different model components on performance with best results in bold.

Model Configuration	ACC (%)	PRE (%)	F1 (%)	AUC (%)
Only Consistent Learning	72.32	71.95	72.13	69.10
W/O Disentanglement Loss	76.96	82.67	77.02	72.48
W/O Multiview Construction	66.96	84.51	76.43	68.96
W/O Multiview and Disentanglement Loss	63.69	65.36	64.02	61.52
Full Model	**86.90**	**93.22**	**90.91**	**89.77**

W/O Disentanglement Loss, (3) **W/O Multiview Construction**, and (4) **W/O Multiview and Disentanglement Loss**. The results demonstrate that relying solely on consistent learning fails to fully leverage the richness of multiview representations, limiting performance. Excluding the disentanglement loss reduces feature separation, increasing redundancy and potentially hindering interpretability. Removing the multiview module weakens the model's ability to capture diverse brain connectivity patterns. Notably, removing both the multiview module and disentanglement loss results in the worst performance, demonstrating the essential and complementary roles of both modules. Overall, the full model's superior performance highlights the importance of integrating multiview construction, masked consistent learning, and disentanglement for effective AD diagnosis.

Table 3. Parameter robustness analysis and the best results are shown in bold.

Configuration	ACC (%)	PRE (%)	F1 (%)	SEN (%)	SPE (%)
$\beta = 1.0$, $\gamma = 1.0$, $\lambda = 1.0$	82.14	82.88	82.40	88.03	**82.45**
$\beta = 0.4$, $\gamma = 0.4$, $\lambda = 0.2$	**86.90**	**93.22**	**90.91**	88.71	81.82
$\beta = 0.7$, $\gamma = 0.2$, $\lambda = 0.1$	82.74	83.06	82.87	82.43	72.45
$\beta = 0.5$, $\gamma = 0.2$, $\lambda = 0.3$	84.52	83.37	84.43	84.43	80.39
$\beta = 0.6$, $\gamma = 0.1$, $\lambda = 0.3$	84.71	85.89	85.79	**88.89**	76.43

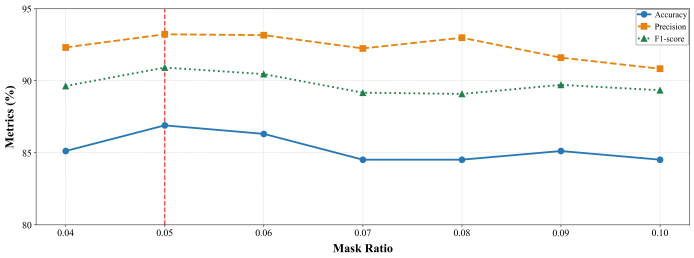

Fig. 2. Effect of mask ratio on model performance (optimal at 0.05).

4 Discussion

Parameter Sensitivity Analysis. *First*, as shown in Fig. 2, model performance peaks at the mask ratio of 0.05, achieving an optimal balance between preserving discriminative information and introducing view diversity. Lower mask ratios provide insufficient diversity, while higher ratios obscure crucial details, degrading performance. *Second*, Table 3 shows the impact of varying weight coefficients (β, γ, λ) in Eq. (2) on model performance. While performance varies slightly across different configurations, the relatively small fluctuations in ACC, PRE, and F1-score indicate the model's robustness to parameter changes.

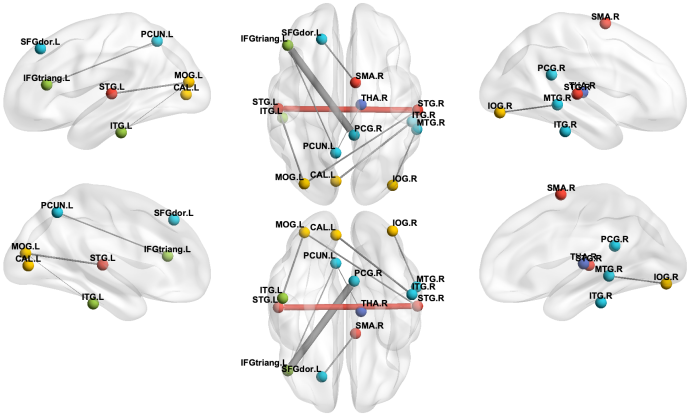

Fig. 3. Illustration of the top ten most discriminative brain connectivities for AD classification, with line thickness indicating connection strength between ROIs.

Discriminative ROIs & Functional Connectivities. As depicted in Fig. 3, we employed the BrainNet Viewer toolbox[2] to visualize the ten most discriminative functional connections, highlighting key brain ROIs and their interactions for AD classification. Notably, connections involving the supplementary motor area (SMA.R), thalamus (THA.R) and superior temporal gyrus (STG.R) highlight their established roles in motor coordination, sensory integration, and cognitive functions. Furthermore, connections including the precuneus (PCUN.L), postcentral gyrus (PCG.R), and inferior occipital gyrus (IOG.R) correspond with reported disruptions in connectivity associated with neurodegenerative conditions, specifically impacting visuospatial processing, memory retrieval, and attentional control in AD [23]. Therefore, these regions represent promising biomarkers for early diagnosis of AD, providing valuable information on the underlying neural mechanisms of the disease.

[2] `https://www.nitrc.org/projects/bnv/`

5 Conclusion and Future Work

In this paper, we propose the **Masked Multiview Brain Network Analysis (MMBNA) Framework** through multiview brain network representation learning for fMRI-based AD diagnosis. Our approach addresses limitations of existing methods through (1) integrating multi-measure brain connectivity estimation with masking to generate topologically diverse yet semantically similar FBN views and (2) disentangling view-specific representations from view-consistent representations, reducing redundancy. Experiments on ADNI2 demonstrate that our framework outperforms SOTA and baseline methods. However, its limitations include the use of the predefined AAL template, potentially introducing biases, and the validation solely on the ADNI2 dataset, restricting generalizability.

Future work will explore adaptive, data-driven brain network construction methods to learn subject-specific ROIs and functional connectivities, overcoming the limitations of fixed anatomical templates. Additionally, we will validate the framework on a wider range of tasks to ensure robustness and broader applicability.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 62176112 and 62476155), and the Natural Science Foundation of Shandong Province (Grant No. ZR2024MF063).

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

References

1. Yu W, Lei B, Wang S, et al. Morphological feature visualization of Alzheimer's disease via multidirectional perception GAN[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 34(8): 4401-4415.
2. Lei B, Liang E, Yang M, et al. Predicting clinical scores for Alzheimer's disease based on joint and deep learning[J]. Expert Systems with Applications, 2022, 187: 115966.
3. Teng Y, Wu K, Liu J, et al. Constructing High-order Functional Connectivity Networks with Temporal Information from fMRI Data[J]. IEEE Transactions on Medical Imaging, 2024.
4. Khosla M, Jamison K, Ngo G H, et al. Machine learning in resting-state fMRI analysis[J]. Magnetic resonance imaging, 2019, 64: 101-121.
5. Wang M, Huang J, Liu M, et al. Modeling dynamic characteristics of brain functional connectivity networks using resting-state functional MRI[J]. Medical image analysis, 2021, 71: 102063.
6. Wang G, Chu Y, Wang Q, et al. Graph Convolutional Network With Self-Supervised Learning for Brain Disease Classification[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2024.
7. Li X, Zhou Y, Dvornek N, et al. Braingnn: Interpretable brain graph neural network for fmri analysis[J]. Medical Image Analysis, 2021, 74: 102233.
8. Wang X, Chu Y, Wang Q, et al. Unsupervised contrastive graph learning for resting-state functional MRI analysis and brain disorder detection[J]. Human Brain Mapping, 2023, 44(17): 5672-5692.

9. Gadgil S, Zhao Q, Pfefferbaum A, et al. Spatio-temporal graph convolution for resting-state fMRI analysis[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23. Springer International Publishing, 2020: 528-538.
10. Kim B H, Ye J C, Kim J J. Learning dynamic graph representation of brain connectome with spatio-temporal attention[J]. Advances in Neural Information Processing Systems, 2021, 34: 4314-4327.
11. Luo G, Li C, Cui H, et al. Multi-view brain network analysis with cross-view missing network generation[C]//2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2022: 108-115.
12. Zhang C, Wang C, Zhang L, et al. Consistent and Specific Multi-View Functional Brain Networks Fusion for Autism Spectrum Disorder Diagnosis[J]. Journal of Applied Mathematics and Physics, 2023, 11(7): 1914-1929.
13. Wen G, Cao P, Bao H, et al. MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis[J]. Computers in biology and medicine, 2022, 142: 105239.
14. Gan J, Peng Z, Zhu X, et al. Brain functional connectivity analysis based on multi-graph fusion[J]. Medical image analysis, 2021, 71: 102057.
15. Zhang C, Ma Y, Qiao L, et al. Learning to fuse multiple brain functional networks for automated autism identification[J]. Biology, 2023, 12(7): 971.
16. Wang C, Zhang L, Zhang J, et al. Fusing Multiview Functional Brain Networks by Joint Embedding for Brain Disease Identification[J]. Journal of Personalized Medicine, 2023, 13(2): 251.
17. Ke G, Wang B, Wang X, et al. Rethinking multi-view representation learning via distilled disentangling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 26774-26783.
18. Lee H, Lee D S, Kang H, et al. Sparse brain network recovery under compressed sensing[J]. IEEE Transactions on Medical Imaging, 2011, 30(5): 1154-1165.
19. Wang Z, Alahmadi A, Zhu D, et al. Brain functional connectivity analysis using mutual information[C]//2015 IEEE global conference on signal and information processing (GlobalSIP). IEEE, 2015: 542-546.
20. Xu K, Hu W, Leskovec J, et al. How powerful are graph neural networks?[J]. arXiv preprint arXiv:1810.00826, 2018.
21. Cheng P, Hao W, Dai S, et al. Club: A contrastive log-ratio upper bound of mutual information[C]//International conference on machine learning. PMLR, 2020: 1779-1788.
22. Chao-Gan Y, Yu-Feng Z D. a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. Front Syst Neurosci. 2010; 4: 13[EB/OL].(2010)
23. Khazaee A, Ebrahimzadeh A, Babajani-Feremi A. Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory[J]. Clinical Neurophysiology, 2015, 126(11): 2132-2141.