HHS Public Access

Author manuscript

IEEE Trans Biomed Eng. Author manuscript; available in PMC 2024 August 01.

Published in final edited form as:

IEEE Trans Biomed Eng. 2024 August ; 71(8): 2391–2401. doi:10.1109/TBME.2024.3370415.

Leveraging Brain Modularity Prior for Interpretable Representation Learning of fMRI

Qianqian Wang,
Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

Wei Wang,
Department of Radiology, Beijing Youan Hospital, Capital Medical University, Beijing 100069, China.

Yuqi Fang,
Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

P.-T. Yap,
Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

Hongtu Zhu,
Department of Biostatistics and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA.

Hong-Jun Li,
Department of Radiology, Beijing Youan Hospital, Capital Medical University, Beijing 100069, China.

Lishan Qiao,
School of Computer Science and Technology, Shandong Jianzhu University, 250101, China.

Mingxia Liu
Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

Abstract

Resting-state functional magnetic resonance imaging (rs-fMRI) can reflect spontaneous neural activities in the brain and is widely used for brain disorder analysis. Previous studies focus on extracting fMRI representations using machine/deep learning methods, but these features typically lack biological interpretability. The human brain exhibits a remarkable modular structure in spontaneous brain functional networks, with each module comprised of functionally interconnected brain regions-of-interest (ROIs). However, existing learning-based methods cannot adequately utilize such brain modularity prior. In this paper, we propose a brain modularity-constrained dynamic representation learning framework for interpretable fMRI analysis, consisting

Corresponding authors: M. Liu (Mingxia_liu@med.unc.edu) and H. Li (lihongjun00113@ccmu.edu.cn).

of dynamic graph construction, dynamic graph learning via a novel modularity-constrained graph neural network (MGNN), and prediction and biomarker detection. The designed MGNN is constrained by three core neurocognitive modules (*i.e.*, salience network, central executive network, and default mode network), encouraging ROIs within the same module to share similar representations. To further enhance discriminative ability of learned features, we encourage the MGNN to preserve network topology of input graphs via a graph topology reconstruction constraint. Experimental results on 534 subjects with rs-fMRI scans from two datasets validate the effectiveness of the proposed method. The identified discriminative brain ROIs and functional connectivities can be regarded as potential fMRI biomarkers to aid in clinical diagnosis.

Keywords

Functional MRI; brain modularity; brain disorder; biomarker

I. Introduction

Resting-state functional magnetic resonance imaging (rs-fMRI) provides a noninvasive solution to reveal brain spontaneous neural activities by measuring blood-oxygenation-level-dependent (BOLD) signals [1]–[3]. It has been increasingly used to understand underlying neuropathological mechanisms of various brain disorders, such as autism spectrum disorder and cognitive impairment [4]–[7]. Many machine/deep learning-based methods have been proposed to map 4D fMRI data into low-dimensional representations and perform downstream brain disease detection [8]–[10]. However, due to the complexity of brain organization and the blackbox property of many learning-based methods, the generated fMRI representations usually lack biological interpretability, thereby limiting their clinical utility [11], [12].

From the perspective of graph theory, the human brain exhibits a significant modular structure in spontaneous brain functional networks (BFN), with each module executing specialized cognitive function [13]. A functional module can be defined as a subnetwork of densely interconnected brain regions-of-interest (ROIs) that are sparsely connected to ROIs in other modules [14], [15]. In particular, salience network (SN), central executive network (CEN), and default mode network (DMN) are three fundamental neurocognitive modules/subnetworks in the human brain, supporting effective cognitive activities [16]. Unfortunately, existing fMRI-based studies usually fail to adequately utilize such brain modularity prior during fMRI representation learning. On the other hand, the brain can be modeled as a spatiotemporally dynamic BFN (*i.e.*, dynamic graph) based on BOLD signals, aiming to help simultaneously capture spatial and temporal information of brain neural activities [17], [18]. Intuitively, it is meaningful to incorporate brain modularity prior into a dynamic graph representation framework for interpretable fMRI analysis.

To this end, we propose a Brain Modularity-constrained dynamic Representation learning (**BMR**) framework for interpretable fMRI analysis. As shown in the top panel of Fig. 1, the proposed BMR consists of three major components: (1) dynamic graph construction using sliding windows, (2) dynamic graph representation learning via a novel brain

modularityconstrained graph neural network (MGNN), and (3) prediction and biomarker detection. Three core neurocognitive modules (*i.e.*, SN, CEN, and DMN) are explicitly incorporated into the BMR through our proposed modularity constraint, encouraging learned features of ROIs within the same module to be similar. To enhance discriminative ability of learned features, we design a graph topology reconstruction constraint to encourage our BMR to preserve the network topology of input graphs during fMRI feature learning. Experimental results on 534 subjects with rs-fMRI from the public Autism Brain Imaging Data Exchange (ABIDE) dataset [19] and a private HIV-associated neurocognitive disorder (HAND) dataset demonstrate the superiority of the proposed BMR over several state-of-the-art methods for brain disorder detection.

The contributions of this work are summarized as follows.

- We propose a brain modularity-constrained dynamic representation learning framework for interpretable fMRI analysis. It provides a novel scheme to incorporate topological prior such as fundamental functional modules in the brain into deep neural networks for fMRI analysis.
- A graph topology reconstruction constraint is introduced to reconstruct the original graph adjacency matrix during graph representation learning, thereby helping extract more discriminative fMRI features.
- The proposed BMR is a general framework for diagnosis of different brain disorders, as evidenced by its superior performance on two fMRI datasets with 534 subjects, when compared with several state-of-the-art methods.

II. Related Work

A. Functional MRI Representation Learning

Various machine learning methods have been used to learn latent representations of resting-state fMRI for brain disorder analysis [20]. For example, Wee *et al.* [21] proposed a constrained sparse linear regression model to estimate brain functional network (BFN) for mild cognitive impairment classification with resting-state fMRI. Rosa *et al.* [22] designed a sparse network-based predictive model that first constructed sparse inverse covariance matrices and then used a sparse support vector machine (SVM) for major depressive disorder detection. Gan *et al.* [23] proposed a multi-graph fusion method that first integrated fully-connected BFNs and one nearest neighbor BFNs and then employed the L1SVM for brain disorder classification. However, these existing studies generally treat fMRI feature learning and downstream model training as two standalone steps, possibly leading to suboptimal performance due to heterogeneity between these steps.

Deep learning methods have been widely used for computer-aided brain disorder diagnosis with fMRI [24], by jointly conducting fMRI representation learning and downstream model training in an end-to-end manner. In particular, due to the graph structure nature of BFN, graph neural networks (GNNs) have shown significant superiority in fMRI representation learning. Ktena *et al.* [25] proposed a siamese graph convolutional network (GCN) to estimate BFN for automated autism analysis. Jiang *et al.* [26] designed a hierarchical

GCN framework for BFN embedding learning by efficiently integrating correlations among subjects in a population. Hu *et al.* [27] designed a complementary graph representation learning method to capture local and global patterns for fMRI-based brain disease analysis. Although these GNN-based methods can model spatial interactions among brain ROIs, they often neglect dynamic variations over time of fMRI data. Considering that temporal dynamics conveyed in fMRI can provide discriminative information for brain disease diagnosis, some GNN-based studies have paid more attention to spatiotemporal dynamic brain network analysis with fMRI data. Gadgil *et al.* [18] introduced a novel spatiotemporal GCN, which captured temporal dynamics within fMRI series via 1D convolutional kernels, to learn dynamic graph representation for age and gender prediction. Even achieving promising results, most existing GNN models cannot explicitly preserve network topology of BFNs during dynamic graph learning.

B. Brain Functional Modularity Analysis

From a graph-theoretic perspective, the BFN during the resting state exhibits a significant modular structure to facilitate efficient information communication and cognitive function [14], [15]. To better understand brain connectivity patterns, researchers have devoted considerable attention to analyzing brain modularity. For example, Meunier *et al.* [28] explored age-related changes in brain modular organization and demonstrated significantly non-random modularity in young and older brain networks. Arnemann *et al.* [29] tested the value of modularity metric to predict response to cognitive training after brain injury. Gallen *et al.* [30] demonstrated that brain modularity could be regarded as a unifying biomarker of intervention-related plasticity by multiple independent studies.

Notably, previous neuroscience studies have demonstrated that there are three fundamental cognitive modules, *i.e.*, salience network (SN), central executive network (CEN), and default mode network (DMN) in human brains. Specifically, SN mainly detects external stimuli and coordinates brain neural resources, CEN performs high-level cognitive tasks (*e.g.*, decision-making and rule-based problem-solving), while DMN is responsible for self-related cognitive functions (*e.g.*, mind-wandering and introspection) [16], [31]. These three modules have been consistently observed across different individuals and experimental paradigms [31], [32]. For example, Menon *et al.* [31] proposed a unifying triple network model comprised of CEN, DMN, and SN, providing a common framework for understanding behavioral and cognitive dysfunction across multiple brain disorders. Krishnadas *et al.* [32] investigated disrupted resting-state functional connectivities within the triple network in patients with paranoid schizophrenia. Intuitively, such brain modularity structures can be employed as important prior knowledge to promote informative fMRI feature learning. However, existing studies typically fail to incorporate such important modularity prior into deep graph learning models for fMRI-based brain disorder analysis.

III. Materials

A. Data Acquisition

A total of 534 subjects with rs-fMRI scans from the public Autism Brain Imaging Data Exchange (ABIDE) dataset [19] and a private HIV-associated neurocognitive disorder

(HAND) dataset are used in this work. On ABIDE, we identify patients with autism spectrum disorder (ASD) from healthy control (HC) subjects on the two largest sites (*i.e.*, NYU and UM). Specifically, the NYU site includes 79 ASDs and 105 HCs, and the UM site includes 68 ASDs and 77 HCs. On HAND, we perform two types of classification tasks, including (1) asymptomatic neurocognitive impairment with HIV (ANI) vs. HC classification, and (2) intact cognition with HIV (ICH) vs. HC classification. Here, rs-fMRI data in the HAND are collected from Beijing Youan Hospital, including 68 ANIs, 68 ICHs and 69 HCs. The demographic information of the studied subjects from two datasets is reported in Table I.

B. Data Preprocessing

All rs-fMRI data from two datasets were preprocessed using the Data Processing Assistant for Resting-State fMRI (DPARSF) pipeline [33]. Specifically, for each fMRI, we first discarded the first 10 time points for magnetization equilibrium. Then, we performed slice timing correction, head motion correction, and regression of nuisance covariates (*e.g.*, white matter signals, ventricle, and head motion parameters). Afterward, the fMRI data were normalized into montreal neurological institute (MNI) space, followed by spatial smoothing with a $4mm$ full-width half maximum Gaussian kernel and band-pass filtering ($0.01 - 0.1\ Hz$). Finally, we extracted the mean rs-fMRI time series of 116 ROIs defined by the automated anatomical labeling (AAL) atlas for each subject.

IV. Proposed Methodology

From the view of graph theory, the brain exhibits a remarkable modular structure in spontaneous brain functional networks [13], [14]. Three fundamental functional modules, including SN, CEN, and DMN, have been consistently observed across different individuals and experimental paradigms [16]. Intuitively, it is meaningful to incorporate such brain modularity prior into a dynamic graph representation framework for interpretable fMRI analysis. Therefore, we propose a cosine similarity-based modularity constraint to encourage brain ROIs within the same functional module to share similar representations. On the other hand, previous fMRI studies have demonstrated that abnormal brain functional networks (BFNs) help reveal underlying pathophysiology of brain disorders [34]. Especially, altered topological structure of BFNs, such as increased global efficiency and decreased local efficiency, can provide potential biomarkers for brain disorder analysis [35], [36]. It is meaningful to ensure that the learned embeddings can capture essential topological information of the original graph during fMRI feature learning. To further enhance discriminative ability of learned fMRI features, a graph topology reconstruction constraint is introduced to preserve the underlying topology of brain FC networks via reconstructing the original graph adjacency matrix during graph representation learning. As illustrated in Fig. 1, the BMR consists of (1) dynamic graph construction using sliding windows, (2) dynamic graph learning via a novel brain modularity-constrained graph neural network (MGNN), and (3) prediction and biomarker detection for interpretable brain disorder analysis. The BMR is an end-to-end trainable model for fMRI-based brain disorder prediction, by jointly learning fMRI features and a predictor.

A. Dynamic Graph Construction

Brain functional network (BFN) derived from fMRI data can capture abnormal connectivity patterns caused by brain disorders by modeling complex dependencies among brain ROIs. Given the fact that brain functional connectivity exhibits dynamic variability over a short period of time [18], we construct a dynamic BFN using sliding windows for each subject. Denote BOLD signals obtained from rs-fMRI as $S \in \mathbb{R}^{N \times M}$, where N is the number of ROIs and M is the number of time points. We first partition the fMRI time series into T segments along the temporal dimension via overlapped sliding windows, with the window size of Γ and the step size of τ. With each ROI treated as a specific node, we construct a fully-connected BFN by calculating Pearson correlation coefficients [37] between segmented fMRI time series of pairwise brain ROIs for each of T segments, obtaining a set of symmetric matrices $\{X_t\}_{t=1}^{T} \in \mathbb{R}^{N \times N}$. Here, the original node feature for the j-th node is represented by the j-th row of X_t for segment t (X_t is also called node feature matrix). Considering that a fully-connected BFN may contain some noisy or redundant information, following [38], we empirically retain the top 30% (*i.e.,* sparsity ratio) strongest edges in each FC network to generate an adjacency matrix $A_t = (a_{ij}) \in \{0,1\}^{N \times N}$ for the segment t, where $a_{ij} = 1$ means there exists an edge between two nodes/ROIs and otherwise $a_{ij} = 0$. Finally, the obtained dynamic graph sequence of each subject can be described as $G_t = \{A_t, X_t\}(t = 1, \cdots, T)$.

B. Dynamic Graph Representation Learning via MGNN

As illustrated in the bottom of Fig. 1, with the constructed dynamic graph sequence $\{G_t\}_{t=1}^{T}$ as input, we design a brain modularity-constrained graph neural network (MGNN) for dynamic fMRI representation learning, including (1) spatial feature learning and (2) temporal feature learning, which can simultaneously model spatial dependencies among brain ROIs and temporal dynamics over time. Notably, a novel *brain modularity constraint* and a *graph topology reconstruction constraint* are incorporated into MGNN to learn more interpretable and discriminative graph representations.

1) Spatial Feature Learning: Considering the graphstructured property of BFNs, we employ a graph attention network (GAT) as the spatial feature encoder to model spatial dynamic representation of BFNs in this work. Taking the segment t as an example, the spatial encoder takes the node feature matrix $X_t = [X_1^t, X_2^t, \cdots, X_N^t]^{\top}\left(X_i^t \in \mathbb{R}^{1 \times N}\right)$ and the graph adjacency/topology matrix A_t as input. Denote $\mathcal{N}_i^t$ as the neighboring node set of the i-th node and $\oplus$ as the concatenation operation. The to-be-learned connection weight (also called spatial attention coefficient) between the i-th ROI and its neighborhood ROI j can be formulated as:

$$\alpha_{ij}^t = \frac{\exp\left(\psi\left(\left[X_i^t W^t \oplus X_j^t W^t\right] \eta^t\right)\right)}{\sum_{v \in \mathcal{N}_i^t} \exp\left(\psi\left(\left[X_i^t W^t \oplus X_v^t W^t\right] \eta^t\right)\right)}, \tag{1}$$

where ψ is a nonlinear activation function (*i.e.*, LeakyRelu), $W^t \in \mathbb{R}^{N \times N'}$ is a shared transformation matrix that maps the original N-dimensional node feature vector to an N'-dimensional vector, and $\eta^t \in \mathbb{R}^{2N'}$ is a to-be-learned weight vector. Then, the updated node representation is expressed as:

$$F_i^t = \sum_{j \in \mathcal{N}_i^t} \alpha_{ij}^t X_j^t W^t, \tag{2}$$

where $F_i^t \in \mathbb{R}^{1 \times N'}$ is the new embedding of the node i after aggregating neighboring node representations. To model different types of spatial dependencies/relationships among ROIs/nodes, we employ a multi-head attention mechanism, which first calculates node representation using multiple attention heads in parallel and then averages them. Mathematically, the output feature $H_i^t \in \mathbb{R}^{1 \times N'}$ of the node i generated by the multi-head attention mechanism can be written as follows:

$$H_i^t = \sigma\left(\frac{1}{K}\sum_{k=1}^{K} F_i^{t,k}\right), \tag{3}$$

where σ is a nonlinear function and K is the number of attention heads. Given N nodes, the new node-level embedding of BFN can be expressed as $H^t = [H_1^t, \cdots, H_N^t]^\top$ at segment t.

(1) Brain Modularity Constraint.: As an important property of BFN [30], modularity provides valuable insights into the organization and integration of brain networks. In general, each module is comprised of densely interconnected brain ROIs that are sparsely connected to ROIs in other modules [14]. Due to the high clustering of connections between ROIs within the module, the brain can locally process specialized cognitive function (*e.g.*, episodic memory processing) with low wiring cost [15]. Previous studies have demonstrated that SN, CEN, and DMN are three fundamental neurocognitive modules in the human brain. Based on such prior knowledge, we reasonably assume that *representations of nodes within the same neurocognitive module tend to be similar to each other*.

Accordingly, we design a unique *brain modularity constraint* during spatial fMRI feature learning in BMR. As illustrated in Fig. 2, the modularity constraint explicitly encourages the nodes belonging to the same module to share similar features. Based on the cosine distance metric, this constraint can be mathematically formulated as follows:

$$L_M = -\sum_{t=1}^{T}\sum_{c=1}^{C}\sum_{i,j=1}^{N_c} \frac{H_i^{t,c} \cdot H_j^{t,c}}{\|H_i^{t,c}\| \cdot \|H_j^{t,c}\|}, \tag{4}$$

where $H_i^{t,c}$ and $H_i^{t,c}$ are representations of node i and node j within the distance metric, this constraint can be mathematically c-th module (with distance metric, this constraint can be mathematically N_c ROIs) at segment t, and C is the number of modules distance metric, this constraint can be mathematically ($C = 3$ in this work). With Eq. (4), we encourage the BMR to focus on brain intrinsic modular organization during fMRI representation learning, thus enhancing discriminative power of learned fMRI features.

(2) Graph Topology Reconstruction Constraint.: To further improve discriminative ability of learned representations, we propose to preserve original topology information of input BFNs by reconstructing the adjacency matrix A_t at segment $t(t = 1, \cdots, T)$. Motivated by previous studies that use variational graph auto-encoders for topology-preserving feature learning [39]–[42], we design a graph decoder in the BMR to predict edge weights between pairwise nodes based on the inner-product of two latent node representations, yielding a reconstructed adjacency matrix $\widehat{A}_t = \sigma\left(H_t \cdot H_t^{\top}\right)$ for segment t, where σ is a nonlinear mapping function. By calculating the inner product of latent node-level representations, it is observed that the reconstructed adjacency matrix $\widehat{A}_t$ is capable of mirroring the cosine similarity among nodes/ROIs based on the learned representation. That is, one can use this reconstructed adjacency matrix to represent the node-level topology. We then propose a *graph topology reconstruction constraint* to help preserve the topology of input BFNs, defined as:

$$L_R = \sum_{t=1}^{T} \xi\left(A_t, \widehat{A}_t\right), \tag{5}$$

where ξ is a cross-entropy loss function. With Eq. 5, we encourage the reconstructed graph to be as similar as possible to the original graph, so the learned node embeddings can capture BFN structure and relationships among ROIs.

To generate graph-level representations, we further apply a squeeze-excitation readout operation [43] based on learned node-level representations. For segment t, the graph-level spatial representation is calculated as:

$$y_t = H_t \Phi\left(P^2 \sigma\left(P^1 H_t \phi_{mean}\right)\right), \tag{6}$$

where Φ is a sigmoid function, P^1 and P^2 are learnable weight matrices, and ϕ_{mean} denotes the average operation.

2) Temporal Feature Learning: As shown in the bottom right of Fig. 1, to further capture temporal dynamics within fMRI series, a single-head transformer encoder is employed to effectively model long-range dependencies across different segments. Here, temporal attention can be measured by a self-attention mechanism in the transformer.

Especially, with spatial graph representation $Y = [y_1, \cdots, y_T]$ as input, the temporal attention matrix can be described as:

$$Z = Softmax\left(\frac{QK^{\top}}{\sqrt{d}}\right), \tag{7}$$

where $Q = \phi_1(Y)$, $K = \phi_2(Y)$, ϕ_1 and ϕ_2 are two linear transformations, d is a scaling factor to stabilize attention mechanism. Then, the graph representation with spatiotemporal attention can be expressed as $Y' = \Psi[Z\phi_3(Y)]$, where $Y' = [y_1', \cdots, y_T']$. ϕ_3 is a linear transformation, and Ψ represents the feed-forward network for further feature abstraction. After that, we sum the updated graph representation sequence $\{y_i'\}_{i=1}^{T}$ to obtain the final whole-graph embedding for subsequent brain disorder analysis.

C. Prediction and Interpretable Biomarker Detection

The whole-graph embedding is finally fed into two fully connected layers and a Softmax layer for brain disease prediction. The objective function of BMR can be formulated as:

$$L = L_C + \lambda_1 L_R + \lambda_2 L_M, \tag{8}$$

where L_C is a cross-entropy loss, L_R and L_M denote the graph topology reconstruction constraint and brain modularity constraint, respectively, while λ_1 and λ_2 are hyperparameters.

For a better interpretation of the graph representations we have learned through BMR, we further delve into the analysis of spatial attention among brain ROIs, aiming at identifying potential biomarkers for supporting brain disorder diagnosis. Specifically, based on spatial attention coefficients described in Eq. (1), we first obtain spatial attention matrices of T segments and average them to generate a spatial attention matrix for each subject. Then, we take the upper triangle elements of each attention matrix, resulting in a 6,670-dimensional vector. Finally, we employ the t-test to select discriminative features by calculating group differences between patients and healthy controls, and map these features to original brain space to detect the most discriminative functional connectivities. The specific biomarker analysis isintroduced in Section VI–B.

D. Implementation Details

The proposed BMR is implemented in PyTorch using a single GPU (NVIDIA TITAN Xp with 12GB memory). The Adam optimizer is used for optimization, with the learning rate of 0.0001, training epochs of 40, batch size of 8, window size of $\Gamma = 40$, and step size of $\tau = 20$. Within the c-th module (with N_c ROIs), we randomly select $m = 50\%$ of all $\frac{N_c(N_c-1)}{2}$ paired ROIs to constrain the BMR. The hyperparameters (*i.e.*, λ_1 and λ_2) in Eq. (8) are determined via a cross-validation (see Section V–A and Section V–E).

V. Experiments

A. Experimental Settings

A 5-fold cross-validation (CV) strategy is employed in the experiments. Within each fold, we randomly select 20% of training samples as the validation set to determine the optimal parameters. We repeat the above 5-fold CV process five times to avoid bias caused by data partition and record the mean and standard deviation results. Six metrics are used to evaluate classification performance, including the area under the receiver operating characteristic curve (AUC), classification accuracy (ACC), F1 score (F1), sensitivity (SEN), specificity (SPE), and balanced accuracy (BAC). A paired sample t-test is used to perform statistical significance analysis between the BMR and each of the competing methods.

B. Competing Methods

We compare the proposed BMR with three machine learning methods: SVM [44], XGBoost [45], and Random Forest [46]; as well as seven deep learning methods: multilayer perceptron (MLP) [47], graph convolutional network (GCN) [48], graph isomorphism network (GIN) [49], graph attention network (GAT) [50], BrainGNN [4], spatio-temporal graph convolutional network (STGCN) [18], and blood-oxygen-level-dependent transformer (BolT) model [51].

1. **SVM:** In this method, we first construct a BFN based on rs-fMRI for each subject by measuring Pearson correlation (PC) coefficients between pairwise brain ROIs. Then, we extract multiple node statistics (*i.e.,* degree centrality, clustering coefficient, betweenness centrality, and eigenvector centrality) of each BFN and concatenate them into a 464-dimensional vector. Finally, the vectorized feature is fed into a linear SVM (with default parameter $C = 1$) for classification.
2. **XGBoost**: Similar to SVM, we first construct a BFN based on PC for each subject and then concatenate the same node statistics into a vectorized representation, followed by XGBoost (with default parameters) for classification.
3. **Random Forest:** This method uses the same fMRI features as SVM and XGBoost, followed by a random forest classifier (with default parameters).
4. **MLP:** This method first extracts node features to represent each subject, and then uses two fully connected layers for feature abstraction and a Softmax layer for brain prediction.
5. **GCN:** In this method, we first construct a BFN using PC for each subject. Two graph convolutional layers are used to update and aggregate node-level representations. Graph-level representations are generated via readout operation, followed by two fully connected layers for classification.
6. **GIN:** This method first constructs a BFN for each subject and then uses two GIN layers with Weisfeiler-Lehman graph isomorphism test for feature learning. We finally obtain graph representations via a readout operation, followed by two fully connected layers and a Softmax layer for classification.

7. **GAT:** Different from GCN and GIN, the GAT uses a graph attention mechanism to learn adaptive edge weights between brain ROIs. In this method, we use two graph attention layers to learn spatial features, a readout operation to generate graph-level vectors, two fully connected layers, and a Softmax layer for classification. Similar to our BMR, the number of attention heads is 4 for each graph attention layer.

8. **BrainGNN:** BrainGNN is specially designed for fMRI analysis, containing an ROI-selection layer for highlighting salient ROIs. With constructed BFNs as input, BrainGNN uses two ROI-aware graph convolutional layers to learn node embeddings, followed by ROI pooling layers. Then, a readout operation is used to convert node-level features into graphlevel representations, followed by two fully connected layers and a Softmax layer for classification.

9. **STGCN:** The STGCN can jointly capture spatial and temporal information of BFNs via spatiotemporal graph convolution (ST-GC) units. It uses two ST-GC units for learning spatiotemporal features with BOLD signals as input. Then, dynamic graph representations are generated via readout operation, followed by two fully connected layers for prediction.

10. **BolT:** The BolT employs a cascade of transformer encoders equipped with a novel fused window attention mechanism for fMRI time series analysis. Specifically, we first extract ROI-level BOLD responses from fMRI data, and feed them to a learnable linear layer, yielding BOLD tokens. After that, we split the time series into temporally-overlapping windows and use a cascade of transformer blocks to process BOLD tokens. For each time window, a separate learnable classification (CLS) token is used within the transformer blocks, where the CLS tokens input to the first block are initialized as tied vectors across windows and eventually become window-specific encoding through transformer blocks. Finally, the learned CLS tokens are averaged across windows, followed by a linear layer for classification.

For a fair comparison, we use the same number of hidden layers (*i.e.*, 2) and the same number of neurons in each hidden layer (*i.e.*, 64) for five GNN-based methods (*i.e.*, GCN, GIN, GAT, BrainGNN, and STGCN). Additionally, we use the same input BFN data for four GNN-based methods (*i.e.*, GCN, GIN, GAT, and BrainGNN), while the STGCN and BolT employ the processed fMRI time series as input.

C. Classification Results

The quantitative results of the proposed BMR and ten competing methods on ABIDE and HAND are reported in Table II and Table III, respectively, where '*' denotes that the proposed BMR is statistically significantly different from a specific competing method via paired sample t-test. From Tables II–III, we have the following interesting observations.

First, our BMR is superior to traditional machine learning methods (*i.e.*, SVM, XGBoost, and Random Forest) by a significant margin on two datasets (*i.e.*, ABIDE and HAND). For example, in terms of AUC values, the BMR yields the improvement of 16.60%, 11.29%, and

12.03% compared with SVM, XGBoost, and Random Forest on the NYU site of ABIDE, respectively. The possible reason is that our BMR can learn informative fMRI representation in an end-to-end manner as needed for downstream tasks compared with these traditional methods that rely on handcrafted node features. *Second*, compared with seven deep models (*i.e.*, MLP, GCN, GIN, GAT, BrainGNN, STGCN, and BolT), our BMR achieves better performance in terms of most metrics on two datasets. For instance, in the task of ANI vs. HC classification on HAND, the BMR improves the AUC value by 4.87%, compared with BrainGNN (a GNN-based model specially designed for brain network analysis). This is probably because our BMR not only focuses more on three inherent functional modules in the brain but also preserves the original topology structure during graph learning, resulting in more discriminative fMRI representation for classification. *Furthermore*, it can be seen that the BMR consistently outperforms STGCN which models short-range temporal dynamics within fMRI via a convolution operation. The possible reason is that BMR can not only capture long-range temporal dependencies within fMRI series via a transformer encoder, but also incorporate crucial modularity prior to the process of dynamic graph representation learning. *Besides*, compared with BolT (a state-of-the-art method for fMRI analysis), BMR achieves better performance and lower standard deviation in most metrics, which further demonstrates the superiority and stability of BMR.

D. Ablation Study

From the theoretical rationale perspective, the brain modularity constraint and the graph topology reconstruction constraint are two complementary constraints in our BMR. Specifically, the design of brain modularity constraint is inspired by the fact that the brain exhibits inherent modular organization in spontaneous brain functional networks. By incorporating brain modularity constraint into graph learning model, we can help the model learn embeddings that are consistent with common neurocognitive subsystems in the brain. On the other hand, given that changes in the topological structure of BFNs are crucial for detecting brain disorders [35], [36], we propose a novel graph topology reconstruction constraint. This constraint is designed to preserve the original graph topology during fMRI feature learning process. In essence, it promotes the reconstruction of the input graph that mirrors the original as closely as possible. This ensures that the node embeddings derived from the learning process effectively reflect the inherent network structure and the interconnections among ROIs.

To quantitatively analyze the necessity of jointly applying these two constraints in Eq. (8), we compare the BMR with its variants: (1) **BMRw/oM** without modularity constraint, (2) **BMRw/oR** without graph topology reconstruction constraint, and (3) **BMRw/oMR** that only uses GAT and Transformer layers for spatiotemporal representation learning, without these two constraints. The experimental results yielded by these methods in ASD vs. HC classification on NYU from ABIDE are reported in Fig. 3. From Fig. 3, we can see that BMR outperforms BMRw/oM without considering the inherent modular structure in the brain. This implies that incorporating brain modularity prior to fMRI representation can help promote performance by learning more discriminative features. Besides, the BMR is superior to BMRw/oR without performing graph topology reconstruction during fMRI representation learning. The underlying reason is that our graph topology reconstruction

constraint helps capture intrinsic spatial information among brain ROIs. In addition, the BMRw/oMR without the proposed two constraints achieves the worst performance in most cases compared with its three counterparts (*i.e.*, BMRw/oM, BMRw/oR, and BMR). The possible reason is that the joint application of these two constraints can help the model focus on modular structures while preserving important connectivity information during graph representation learning.

E. Influence of Hyperparameters

We have two hyperparameters (*i.e.*, λ_1 and λ_2) in the proposed BMR (see Eq. (8)) to control contributions of brain modularity constraint and graph topology reconstruction constraint, respectively. To study their influences on the performance of BMR, we tune λ_1 and λ_2 within the range of $\{10^{-4},10^{-3}, \cdots, 10^{1}\}$ based on training and validation sets for ASD vs. HC classification on NYU from ABIDE, and report the results of BMR in Fig. 4. It can be observed from Fig. 4 that the BMR with a large λ_1 (*e.g.*, $\lambda_1 = 10$) achieves worse performance. The underlying reason may be that using a strong graph reconstruction constraint will make the model difficult to converge, thus degrading its learning performance. On the other hand, the BMR with a very weak modularity constraint (*e.g.*, $\lambda_2 = 10^{-4}$) is generally inferior to that with relatively stronger modularity constraint (*e.g.*, $\lambda_2 = 10^{-2}$). These results suggest that the BMR can not achieve satisfactory performance when the BMR pays less attention to the brain's inherent modular structure, which further validates the effectiveness of the designed brain modularity constraint. In particular, the BMR achieves the best AUC values with $\lambda_1 = 10^{-2}$ and $\lambda_2 = 10^{-2}$ in this task.

F. Influence of Spatial Feature Encoder

In the main experiments, our BMR uses GAT as spatial feature encoder to capture dependencies among brain ROIs. To investigate the influence of the spatial feature encoder, we replace GAT with the graph isomorphism network (GIN) to extract spatial fMRI features in BMR, and call this variant as **BMR-GIN.** The results of BMR and BMR-GIN for ASD vs. HC classification on NYU are reported in Fig. 5. It can be found from Fig. 5 that BMR achieves better performance than BMR-GIN in most cases. The main reason could be that, compared with BMR-GIN that treats neighboring nodes equally during the process of aggregating node features, the BMR can adaptively assign different attention weights to different neighboring nodes so that the model can focus on important nodes, thus improving learning performance.

G. Influence of Modularity Ratio

In the proposed modularity constraint, we randomly select $m = 50\%$ of all $\frac{N_c(N_c-1)}{2}$ paired ROIs in the c-th module (with N_c ROIs) to constrain the BMR. To explore the influence of modularity ratio, we vary its values within the range of {0%,25%, ⋯, 100%}, and report the results on NYU site of ABIDE in Fig. 6. As shown in Fig. 6, with $m < 75\%$, the ACC and AUC values of BMR generally improve as the increase of m. With a very large modularity ratio (*e.g.*, $m = 100\%$), BMR cannot achieve satisfactory performance. The possible reason

is that using a too strong modularity constraint in BMR may lead to an over-smoothing problem, thus weakening the discriminative power of learned representations.

H. Influence of Sliding Window Size

In the main experiments, we use the sliding window strategy to generate dynamic BFNs with the window size of $\Gamma = 40$. To further explore the influence of sliding window size, we vary sliding window size within {30,40, ⋯, 80} and record the results of BMR on NYU site of ABIDE in Fig. 7. As shown in Fig. 7, the BMR consistently yields promising performance (*i.e.*, AUC > 72%) when the window size is within the range (*i.e.*, $30 \leq \Gamma \leq 60$). But with large size of sliding windows (*e.g.*, $\Gamma = 80$), BMR cannot achieve good performance. The reason could be that a larger window size provides lower temporal resolution, so the BMR can not effectively capture temporal fluctuations within fMRI series.

I. Influence of Distance Metric

We employ the cosine distance in BMR to quantify the similarity between latent node-level representations within each module, as shown in Eq. 4. To study the effect of this metric, we compare BMR with its three variants: (1) **BMR-ED** with Euclidean distance, (2) **BMR-HD** with Hamming distance, and (3) **BMR-JD** with Jaccard similarity, with results reported in Table IV. As shown in Table IV, BMR using four different distance metrics in the modularity constraint achieves comparable results. This implies that our BMR is not sensitive to the distance metrics used in the modularity constraint.

J. Influence of BFN Sparsity Ratio

Following [38], we empirically retain the top 30% strongest edges (*i.e.*, sparsity ratio) in each BFN in the experiments. To study the impact of sparsity ratio, we vary its value within {10%, ⋯, 100%} and report AUC and ACC values in ASD vs. HC classification on NYU in Fig. 8. As shown in Fig. 8, our BMR achieves stable results when the sparsity ratio is < 60%. For example, BMR obtains the AUC values of 73.69% and 73.32% when sparsity ratios are set as 10% and 50%, respectively. The reason could be that BMR retains the most reliable and informative connections in BFNs by prioritizing the strongest edges, reducing the impact of noisy or redundant connections. But when the sparsity rate is large (*e.g.*, > 90%), the BMR cannot produce good results. The possible reason is that the BFN with such large sparsity can not effectively reflect topology information due to the loss of too many connections.

VI. Discussion

A. Comparison with Previous Studies

In this paper, we develop a novel BMR framework for brain disorder analysis with fMRI data. Compared with previous studies for fMRI analysis [8], [52], [53], our method simultaneously considers brain inherent modular structure and original graph topology information during dynamic graph representation. Extensive experiments on different datasets validate the superiority of our BMR in brain disorder diagnosis.

In previous studies, researchers have demonstrated that there are three fundamental functional modules (*i.e.*, SN, CEN, andDMN) in our brain to support efficient cognition [16]. On the other hand, graph neural networks (GNNs) have been widely used for fMRI-based brain disorder analysis thanks to powerful graph representation ability [25]–[27]. For example, Azevedo [8] proposed a deep neural network architecture that combined both GNNs and temporal convolutional networks to learn both the spatial and temporal features of rs-fMRI data. However, previous works neglect to integrate important modularity prior into GNN-based graph learning models for fMRI-based brain disorder diagnosis, limiting model performance. To this end, we propose the BMR method, where a brain modularity constraint and a graph topology reconstruction constraint are designed to enhance discriminative ability of learned fMRI features. As shown in Table II–III, our BMR achieves better classification results than competing methods on different datasets, which further validates its effectiveness.

B. Discriminative Brain ROI and Functional Connectivity

We also visualize the top 10 discriminative functional connectivities (FCs) identified by the BMR on different datasets (*i.e.*, ABIDE and HAND) in Fig. 9. The thickness of each line represents discriminative ability of the corresponding FC (inversely proportional to the p-value obtained by t-test). For ASD identification (see Fig. 9 (a)), the most discriminative FCs involve *anterior cingulate and paracingulate gyri*, *parahippocampal gyrus*, and *hippocampus*, which complies with previous ASD-related findings [54]–[56]. As shown in Fig. 9 (b), the discriminative brain ROIs in ANI identification include *insula*, *right temporal pole: superior temporal gyrus*, *supplementary motor area*, and *caudate nucleus*. These regions have also been reported in previous studies on HIV-related cognitive impairment [57]–[60]. These results further demonstrate the effectiveness of the BMR in detecting interpretable disease-associated biomarkers. Furthermore, we provide the top 10 most discriminative FCs, related ROIs, and corresponding p-values in Table V. It can be found from Table V that the corresponding p-values are < 0.05 for the top ten FCs, which suggests the strong discriminative ability of identified FCs.

C. Limitations and Future Work

Several limitations need to be considered in future work. *First*, we only characterize pairwise relationships of ROIs within three prominent neurocognitive modules (*i.e.*, SN, CEN, and DMN) as prior knowledge to design the modularity constraint in BMR. It is meaningful to design disease-specific modularity constraints based on neurocognitive research and clinical experience in the future. *Second*, the BMR incorporates known brain modular organization into fMRI representation learning. Future work will seek to design new algorithms that can automatically detect unknown brain modular structures during graph/BFN learning to characterize disease-induced brain changes. *Besides*, the BMR needs to be trained on labeled fMRI data in a supervised manner. In future work, we will employ unsupervised contrastive learning strategies [61] to pre-train the feature encoder on large-scale unlabeled data to learn more discriminative fMRI features.

VII. Conclusion

In this paper, we propose a Brain Modularity-constrained dynamic Representation learning (BMR) framework for interpretable fMRI analysis. Specifically, we first construct a dynamic graph/BFN for each subject, and then design a brain modularity-constrained GNN model for dynamic graph representation learning, where a novel modularity constraint is developed to encourage nodes within the same module to share similar embeddings. We also propose a graph topology reconstruction constraint to preserve original topology information of input BFNs during representation learning. Finally, we perform brain disorder prediction and biomarker detection by analyzing disease-related functional connectivities and brain regions, aiming to provide biological evidence for clinical practice. Extensive experiments demonstrate the effectiveness of BMR in fMRI-based brain disorder detection.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

We thank Mr. Mengqi Wu for his valuable help in writing this paper, and acknowledge the assistance of Yuanyuan Wang, Yu Qi, Shuai Han, Xire Aili, and Yuxun Gao who helped with imaging data collection in the HIV-Associated Neurocognitive Disorder (HAND) study. Y. Fang, P.-T. Yap, and M. Liu were supported by NIH grant AG073297. P.-T. Yap was supported by NIH grant EB035160. The research of M. Liu and H. Zhu was partially supported by NIH grant RF1AG082938. Q. Wang and W. Wang contribute equally to this work.

References

[1]. Bolt T, Nomi JS, Bzdok D, Salas JA, Chang C, Thomas Yeo B, Uddin LQ, and Keilholz SD, “A parsimonious description of global functional brain organization in three spatiotemporal patterns,” Nature Neuroscience, vol. 25, no. 8, pp. 1093–1103, 2022. [PubMed: 35902649]

[2]. Pervaiz U, Vidaurre D, Gohil C, Smith SM, and Woolrich MW, “Multi-dynamic modelling reveals strongly time-varying resting fMRI correlations,” Medical Image Analysis, vol. 77, p. 102366, 2022.

[3]. Sip V, Hashemi M, Dickscheid T, Amunts K, Petkoski S, and Jirsa V, “Characterization of regional differences in resting-state fMRI with a data-driven network model of brain dynamics,” Science Advances, vol. 9, no. 11, p. eabq7547, 2023.

[4]. Li X, Zhou Y, Dvornek N, Zhang M, Gao S, Zhuang J, Scheinost D, Staib LH, Ventola P, and Duncan JS, “BrainGNN: Interpretable brain graph neural network for fMRI analysis,” Medical Image Analysis, vol. 74, p. 102233, 2021.

[5]. Kunda M, Zhou S, Gong G, and Lu H, “Improving multi-site autism classification via site-dependence minimization and second-order functional connectivity,” IEEE Transactions on Medical Imaging, vol. 42, no. 1, pp. 55–65, 2022. [PubMed: 36054402]

[6]. DSouza AM, Abidin AZ, Schifitto G, and Wismuller A, “Ä multivoxel pattern analysis framework with mutual connectivity analysis investigating changes in resting state connectivity in patients with HIV associated neurocognitve disorder,” Magnetic Resonance Imaging, vol. 62, pp. 121–128, 2019. [PubMed: 31189074]

[7]. Wang N, Yao D, Ma L, and Liu M, “Multi-site clustering and nested feature extraction for identifying autism spectrum disorder with restingstate fMRI,” Medical Image Analysis, vol. 75, p. 102279, 2022.

[8]. Azevedo T, Campbell A, Romero-Garcia R, Passamonti L, Bethlehem RA, Lio P, and Toschi N, “A deep graph neural network architecture for modelling spatio-temporal dynamics in resting-state functional MRI data,” Medical Image Analysis, vol. 79, p. 102471, 2022.

[9]. Bessadok A, Mahjoub MA, and Rekik I, “Graph neural networks in network neuroscience,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 5, pp. 5833–5848, 2022.

[10]. Fang Y, Wang M, Potter GG, and Liu M, “Unsupervised cross-domain functional MRI adaptation for automated major depressive disorder identification,” Medical Image Analysis, vol. 84, p. 102707, 2023.

[11]. Geirhos R, Jacobsen J-H, Michaelis C, Zemel R, Brendel W, Bethge M, and Wichmann FA, “Shortcut learning in deep neural networks,” Nature Machine Intelligence, vol. 2, no. 11, pp. 665–673, 2020.

[12]. Chen H, Gomez C, Huang C-M, and Unberath M, “Explainable medical imaging AI needs human-centered design: Guidelines and evidence from a systematic review,” NPJ Digital Medicine, vol. 5, no. 1, p. 156, 2022. [PubMed: 36261476]

[13]. Liao X, Cao M, Xia M, and He Y, “Individual differences and time-varying features of modular brain architecture,” NeuroImage, vol. 152, pp. 94–107, 2017. [PubMed: 28242315]

[14]. Sporns O. and Betzel RF, “Modular brain networks,” Annual Review of Psychology, vol. 67, pp. 613–640, 2016.

[15]. Meunier D, Lambiotte R, and Bullmore ET, “Modular and hierarchically modular organization of brain networks,” Frontiers in Neuroscience, vol. 4, p. 200, 2010. [PubMed: 21151783]

[16]. Goulden N, Khusnulina A, Davis NJ, Bracewell RM, Bokde AL, McNulty JP, and Mullins PG, “The salience network is responsible for switching between the default mode network and the central executive network: Replication from DCM,” NeuroImage, vol. 99, pp. 180–190, 2014. [PubMed: 24862074]

[17]. Kong Y, Gao S, Yue Y, Hou Z, Shu H, Xie C, Zhang Z, and Yuan Y, “Spatio-temporal graph convolutional network for diagnosis and treatment response prediction of major depressive disorder from functional connectivity,” Human Brain Mapping, vol. 42, no. 12, pp. 3922–3933, 2021. [PubMed: 33969930]

[18]. Gadgil S, Zhao Q, Pfefferbaum A, Sullivan EV, Adeli E, and Pohl KM, “Spatio-temporal graph convolution for resting-state fMRI analysis,” in International Conference on Medical Image Computing and Computer Assisted Intervention. Springer, 2020, pp. 528–538.

[19]. Di Martino A, Yan C-G, Li Q, Denio E, Castellanos FX, Alaerts K, Anderson JS, Assaf M, Bookheimer SY, Dapretto M. et al. , “The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism,” Molecular Psychiatry, vol. 19, no. 6, pp. 659–667, 2014. [PubMed: 23774715]

[20]. Khosla M, Jamison K, Ngo GH, Kuceyeski A, and Sabuncu MR, “Machine learning in resting-state fMRI analysis,” Magnetic Resonance Imaging, vol. 64, pp. 101–121, 2019. [PubMed: 31173849]

[21]. Wee C-Y, Yap P-T, Zhang D, Wang L, and Shen D, “Constrained sparse functional connectivity networks for MCI classification,” in International Conference on Medical Image Computing and ComputerAssisted Intervention. Springer, 2012, pp. 212–219.

[22]. Rosa MJ, Portugal L, Hahn T, Fallgatter AJ, Garrido MI, Shawe-Taylor J, and Mourao-Miranda J, “Sparse network-based models for patient classification using fMRI,” NeuroImage, vol. 105, pp. 493–506, 2015. [PubMed: 25463459]

[23]. Gan J, Peng Z, Zhu X, Hu R, Ma J, and Wu G, “Brain functional connectivity analysis based on multi-graph fusion,” Medical Image Analysis, vol. 71, p. 102057, 2021.

[24]. Yin W, Li L, and Wu F-X, “Deep learning for brain disorder diagnosis based on fMRI images,” Neurocomputing, vol. 469, pp. 332–345, 2022.

[25]. Ktena SI, Parisot S, Ferrante E, Rajchl M, Lee M, Glocker B, and Rueckert D, “Metric learning with spectral graph convolutions on brain connectivity networks,” NeuroImage, vol. 169, pp. 431–442, 2018. [PubMed: 29278772]

[26]. Jiang H, Cao P, Xu M, Yang J, and Zaiane O, “Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction,” Computers in Biology and Medicine, vol. 127, p. 104096, 2020.

[27]. Hu R, Peng L, Gan J, Shi X, and Zhu X, “Complementary graph representation learning for functional neuroimaging identification,” in Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 3385–3393.

[28]. Meunier D, Achard S, Morcom A, and Bullmore E, “Age-related changes in modular organization of human brain functional networks,” NeuroImage, vol. 44, no. 3, pp. 715–723, 2009. [PubMed: 19027073]

[29]. Arnemann KL, Chen AJ-W, Novakovic-Agopian T, Gratton C, Nomura EM, and D’Esposito M, “Functional brain network modularity predicts response to cognitive training after brain injury,” Neurology, vol. 84, no. 15, pp. 1568–1574, 2015. [PubMed: 25788557]

[30]. Gallen CL and D’Esposito M, “Brain modularity: A biomarker of intervention-related plasticity,” Trends in Cognitive Sciences, vol. 23, no. 4, pp. 293–304, 2019. [PubMed: 30827796]

[31]. Menon V, “Large-scale brain networks and psychopathology: A unifying triple network model,” Trends in Cognitive Sciences, vol. 15, no. 10, pp. 483–506, 2011. [PubMed: 21908230]

[32]. Krishnadas R, Ryali S, Chen T, Uddin L, Supekar K, Palaniyappan L, and Menon V, “Resting state functional hyperconnectivity within a triple network model in paranoid schizophrenia,” The Lancet, vol. 383, p. S65, 2014.

[33]. Yan C. and Zang Y, “DPARSF: A MATLAB toolbox for “pipeline” data analysis of resting-state fMRI,” Frontiers in Systems Neuroscience, vol. 4, p. 13, 2010. [PubMed: 20577591]

[34]. Woodward ND and Cascio CJ, “Resting-state functional connectivity in psychiatric disorders,” JAMA Psychiatry, vol. 72, no. 8, pp. 743–744, 2015. [PubMed: 26061674]

[35]. Woo C-W, Chang LJ, Lindquist MA, and Wager TD, “Building better biomarkers: Brain models in translational neuroimaging,” Nature Neuroscience, vol. 20, no. 3, pp. 365–377, 2017. [PubMed: 28230847]

[36]. Yang H, Chen X, Chen Z-B, Li L, Li X-Y, Castellanos FX, Bai T-J, Bo Q-J, Cao J, Chang Z-K et al. , “Disrupted intrinsic functional brain topology in patients with major depressive disorder,” Molecular Psychiatry, vol. 26, no. 12, pp. 7363–7371, 2021. [PubMed: 34385597]

[37]. Zhang S, Chen X, Shen X, Ren B, Yu Z, Yang H, Jiang X, Shen D, Zhou Y, and Zhang X-Y, “A-GCL: Adversarial graph contrastive learning for fMRI analysis to diagnose neurodevelopmental disorders,” Medical Image Analysis, vol. 90, p. 102932, 2023.

[38]. Kim B-H, Ye JC, and Kim J-J, “Learning dynamic graph representation of brain connectome with spatio-temporal attention,” Advances in Neural Information Processing Systems, vol. 34, pp. 4314–4327, 2021.

[39]. Kipf TN and Welling M, “Variational graph auto-encoders,” arXiv preprint arXiv:1611.07308, 2016.

[40]. Zhang H, Li P, Zhang R, and Li X, “Embedding graph auto-encoder for graph clustering,” IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 11, pp. 9352–9362, 2022.

[41]. Cheng J, Wang Q, Tao Z, Xie D, and Gao Q, “Multi-view attribute graph convolution networks for clustering,” in Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, 2021, pp. 2973–2979.

[42]. Noman F, Ting C-M, Kang H, Phan RC-W, and Ombao H, “Graph autoencoders for embedding learning in brain networks and major depressive disorder identification,” IEEE Journal of Biomedical and Health Informatics, 2024.

[43]. Hu J, Shen L, and Sun G, “Squeeze-and-excitation networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[44]. Noble WS, “What is a support vector machine?” Nature Biotechnology, vol. 24, no. 12, pp. 1565–1567, 2006.

[45]. Chen T. and Guestrin C, “XGBoost: A scalable tree boosting system,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[46]. Biau G. and Scornet E, “A random forest guided tour,” Test, vol. 25, pp. 197–227, 2016.

[47]. Tang J, Deng C, and Huang G-B, “Extreme learning machine for multilayer perceptron,” IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 4, pp. 809–821, 2015. [PubMed: 25966483]

[48]. Kipf TN and Welling M, “Semi-supervised classification with graph convolutional networks,” arXiv preprint arXiv:1609.02907, 2016.

[49]. Kim B-H and Ye JC, “Understanding graph isomorphism network for rs-fMRI functional connectivity analysis,” Frontiers in Neuroscience, p. 630, 2020. [PubMed: 32714130]

[50]. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, and Bengio Y, “Graph attention networks,” arXiv preprint arXiv:1710.10903, 2017.

[51]. Bedel HA, Sivgin I, Dalmaz O, Dar SU, and T. Çukur, “BolT: Fused window transformers for fMRI time series analysis,” Medical Image Analysis, vol. 88, p. 102841, 2023.

[52]. Zhang H, Song R, Wang L, Zhang L, Wang D, Wang C, and Zhang W, “Classification of brain disorders in rs-fMRI via local-toglobal graph neural networks,” IEEE Transactions on Medical Imaging, vol. 42, no. 2, pp. 444–455, 2022.

[53]. Cui H, Dai W, Zhu Y, Kan X, Gu AAC, Lukemire J, Zhan L, He L, Guo Y, and Yang C, “BrainGB: A benchmark for brain network analysis with graph neural networks,” IEEE Transactions on Medical Imaging, vol. 42, no. 2, pp. 493–506, 2022.

[54]. Dichter GS, Felder JN, and Bodfish JW, “Autism is characterized by dorsal anterior cingulate hyperactivation during social target detection,” Social Cognitive and Affective Neuroscience, vol. 4, no. 3, pp. 215–226, 2009. [PubMed: 19574440]

[55]. Monk CS, Peltier SJ, Wiggins JL, Weng S-J, Carrasco M, Risi S, and Lord C, “Abnormalities of intrinsic functional connectivity in autism spectrum disorders,” NeuroImage, vol. 47, no. 2, pp. 764–772, 2009. [PubMed: 19409498]

[56]. Banker SM, Gu X, Schiller D, and Foss-Feig JH, “Hippocampal contributions to social and cognitive deficits in autism spectrum disorder,” Trends in Neurosciences, vol. 44, no. 10, pp. 793–807, 2021. [PubMed: 34521563]

[57]. Zhou Y, Li R, Wang X, Miao H, Wei Y, Ali R, Qiu B, and Li H, “Motor-related brain abnormalities in HIV-infected patients: A multimodal MRI study,” Neuroradiology, vol. 59, no. 11, pp. 1133–1142, 2017. [PubMed: 28889255]

[58]. Zhan Y, Yu Q, Cai D-C, Ford JC, Shi X, Fellows AM, Clavier OH, Soli SD, Fan M, Lu H. et al. , “The resting state central auditory network: A potential marker of HIV-related central nervous system alterations,” Ear and Hearing, vol. 43, no. 4, p. 1222, 2022. [PubMed: 35044995]

[59]. Shin N-Y, Hong J, Choi JY, Lee S-K, Lim SM, and Yoon U, “Retrosplenial cortical thinning as a possible major contributor for cognitive impairment in HIV patients,” European Radiology, vol. 27, pp. 4721–4729, 2017. [PubMed: 28409354]

[60]. Chockanathan U, DSouza AM, Abidin AZ, Schifitto G, and Wismuller A, “Automated diagnosis of HIV-associated neurocognitivë disorders using large-scale granger causality analysis of resting-state functional MRI,” Computers in Biology and Medicine, vol. 106, pp. 24–30, 2019. [PubMed: 30665138]

[61]. Wang X. and Qi G-J, “Contrastive learning with stronger augmentations,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 5, pp. 5549–5560, 2022.

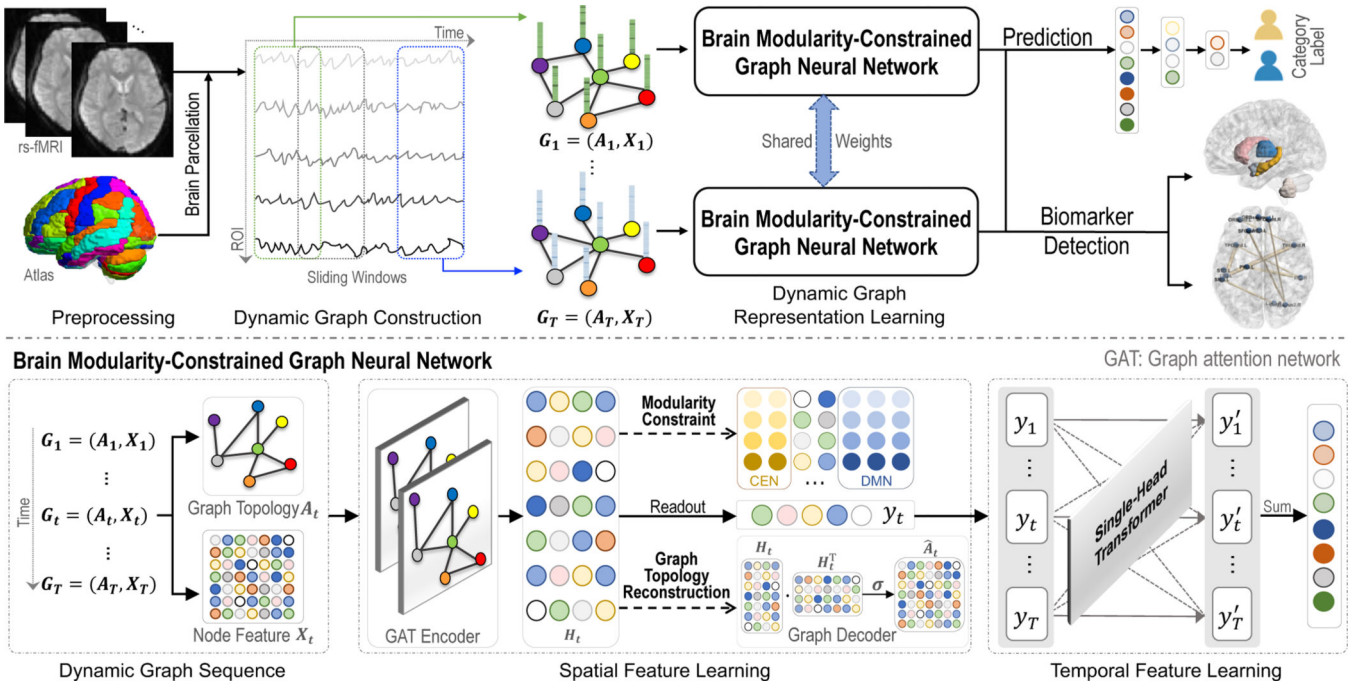

Fig. 1.

Illustration of the proposed Brain Modularity-constrained dynamic Representation learning (BMR) framework, including (1) dynamic graph construction using sliding windows, (2) dynamic graph representation learning via a novel modularity-constrained graph neural network (MGNN), and (3) prediction and biomarker detection for brain disorder analysis.

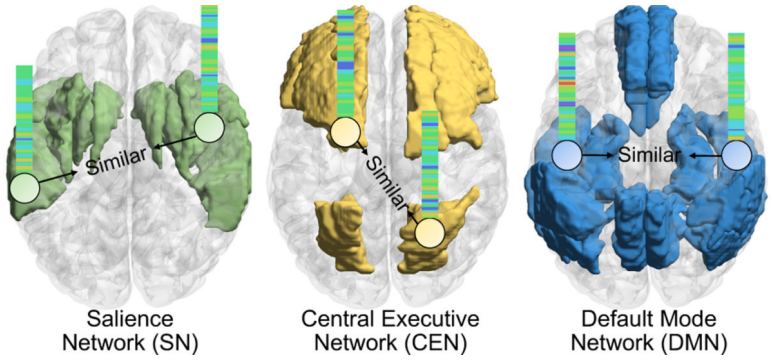

Fig. 2.

Illustration of the proposed modularity constraint with three fundamental cognitive modules, *i.e.*, salience network (SN), central executive network (CEN), and default mode network (DMN), where nodes within the same module are encouraged to share similar representation.

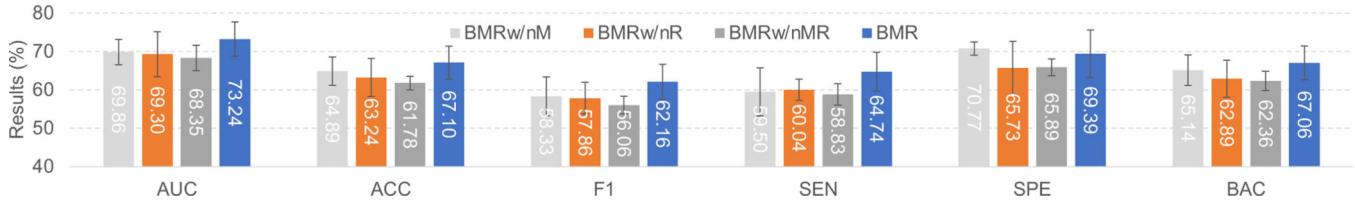

Fig. 3.

Performance of the BMR and its three variants in ASD vs. HC classification on NYU site of ABIDE dataset.

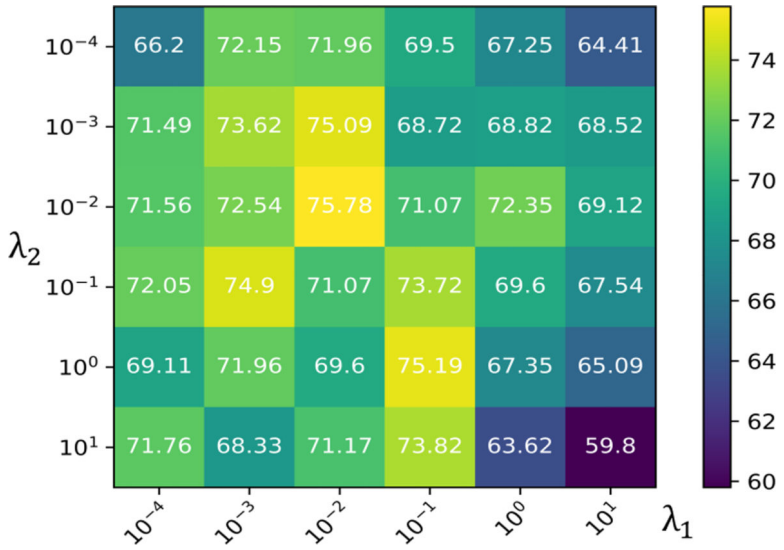

Fig. 4.

AUC values (%) of the BMR under different hyperparameters (*i.e.*, λ_1 and λ_2) in ASD vs. HC classification on NYU of ABIDE.

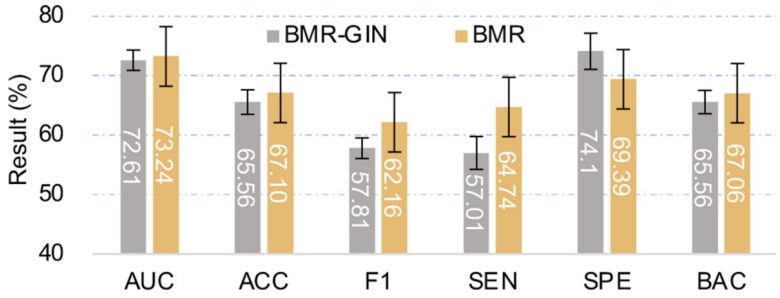

Fig. 5.

Results of our BMR and its variant BMR-GIN (with GIN as spatial encoder) in ASD vs. HC classification on NYU.

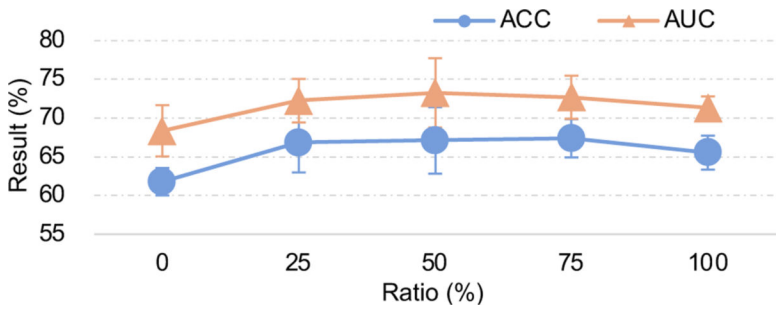

Fig. 6.

Results of the proposed BMR with different modularity ratios in ASD vs. HC classification on NYU.

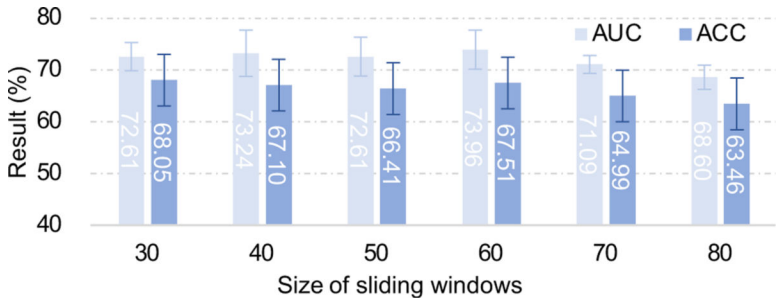

Fig. 7.

Results of our BMR using different sizes of sliding windows in ASD vs. HC classification on NYU.

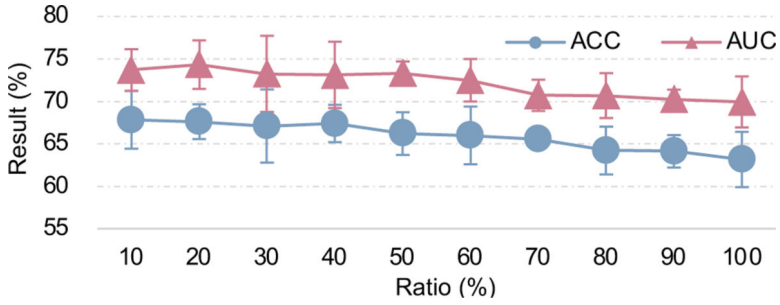

Fig. 8.

Results of the proposed BMR with different sparsity ratios of BFN in ASD vs. HC classification on NYU.

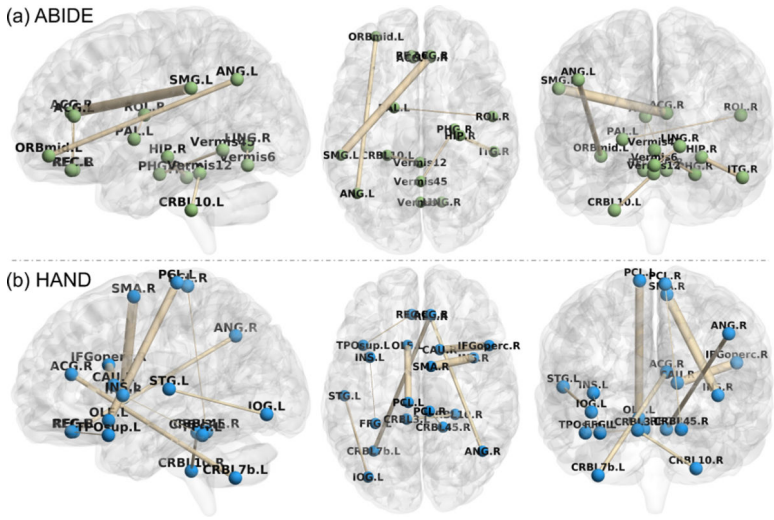

Fig. 9.

Top ten discriminative functional connectivities identified by our BMR in (a) ASD vs. HC classification on NYU from the ABIDE dataset and (b) ANI vs. HC classification on the HAND dataset.

TABLE I

DEMOGRAPHIC INFORMATION OF SUBJECTS FROM TWO LARGEST SITES (*i.e.*, NYU AND UM) of ABIDE DATASET AND A PRIVATE HAND DATASET. ASD: AUTISM SPECTRUM DISORDER; HC: HEALTHY CONTROL; ANI: ASYMPTOMATIC NEUROCOGNITIVE IMPAIRMENT WITH HIV; ICH: INTACT COGNITION WITH HIV; M: MALE; F: FEMALE; STD: STANDARD DEVIATION.

Dataset	Site	Category	Subject #	Gender (M/F)	Age (Mean±Std)
ABIDE	NYU	ASD	79	68/11	14.52 ± 6.97
		HC	105	79/26	15.81 ± 6.25
	UM	ASD	68	58/10	13.13 ± 2.41
		HC	77	59/18	14.79 ± 3.57
HAND	-	ANI	68	68/0	33.07 ± 6.18
		ICH	68	68/0	33.40 ± 5.58
		HC	69	69/0	33.33 ± 5.37

TABLE II

CLASSIFICATION RESULTS OF COMPETING METHODS AND THE PROPOSED BMR ON THE TWO LARGEST SITES OF ABIDE. RESULTS ARE SHOWN IN THE FORM OF "MEAN(STANDARD DEVIATION)" AND THE BEST RESULTS ARE SHOWN IN BOLD.

Method	ASD vs. HC classification on NYU						ASD vs. HC classification on UM					
	AUC (%)	ACC (%)	FI (%)	SEN (%)	SPE (%)	BAC (%)	AUC (%)	ACC (%)	FI (%)	SEN (%)	SPE (%)	BAC (%)
SVM	56.64(2.89)*	54.83(3.11)	48.64(3.38)	51.48(4.56)	57.88(4.66)	54.69(3.09)	53.61(4.25)*	53.32(3.63)	49.32(4.20)	50.29(4.53)	56.60(4.66)	53.45(3.76)
XGBoost	61.95(0.56)*	63.00(1.63)	51.49(1.84)	47.99(2.68)	75.91(3.72)	61.95(0.56)	58.81(0.83)*	58.62(1.89)	50.19(1.95)	47.61(3.64)	70.01(3.94)	58.81(0.84)
Random Forest	61.21(3.09)*	61.10(4.48)	49.31(4.62)	46.04(4.12)	74.06(5.23)	60.05(4.67)	57.25(2.77)*	56.13(3.66)	49.25(4.87)	47.73(5.83)	65.68(4.11)	56.70(4.04)
MLP	58.64(2.51)*	58.28(1.60)	46.66(4.39)	45.22(7.22)	68.24(5.48)	56.73(1.64)	58.17(2.89)*	54.48(3.00)	48.06(1.81)	47.67(2.47)	62.70(5.99)	55.19(2.78)
GCN	67.53(3.34)	63.59(3.05)	53.63(3.75)	50.99(5.09)	73.54(4.69)	62.26(2.82)	66.74(2.58)*	60.00(2.96)	55.30(3.07)	54.61(4.42)	66.50(5.22)	60.56(2.57)
GIN	61.43(3.45)*	57.04(2.48)	48.24(2.13)	49.35(5.39)	64.94(5.70)	57.14(1.16)	58.94(3.19)*	56.89(2.73)	50.47(2.73)	49.62(3.23)	64.92(3.41)	57.27(2.40)
GAT	64.87(2.64)*	60.12(2.64)	52.14(3.83)	52.96(4.92)	66.12(3.15)	59.54(2.79)	67.34(3.26)*	60.90(3.58)	55.54(4.59)	54.98(5.39)	68.21(5.15)	61.60(2.95)
BrainGNN	66.89(2.90)*	63.21(3.15)	56.31(4.17)	57.12(4.81)	68.51(3.03)	62.82(3.24)	65.91(2.47)	62.69(2.55)	57.18(1.21)	55.47(3.25)	68.09(5.67)	61.78(2.06)
STGCN	66.69(0.87)*	61.54(1.65)	45.55(2.23)	53.64(2.34)	68.42(1.75)	61.03(1.53)	64.01(0.14)*	63.90(0.10)	44.19(3.91)	55.95(1.18)	72.16(0.56)	64.07(0.14)
BolT	69.94(7.45)*	65.01(5.11)	49.54(9.86)	42.28(10.61)	**82.10(6.26)**	62.19(5.54)	69.90(8.42)	61.24(4.58)	49.89(8.11)	42.54(11.60)	**78.22(12.01)**	60.38(4.63)
BMR (Ours)	**73.24(4.48)**	**67.10(4.29)**	**62.16(4.53)**	**64.74(5.08)**	69.39(6.22)	**67.06(4.38)**	**70.55(5.22)**	**65.28(2.37)**	**62.25(2.14)**	**63.21(2.71)**	67.42(4.79)	**65.31(3.08)**

'*' THE TERM REPRESENTS THE PROPOSED BMR IS STATISTICALLY SIGNIFICANTLY DIFFERENT FROM A COMPETING METHOD.

IEEE Trans Biomed Eng. Author manuscript; available in PMC 2024 August 01.

TABLE III

CLASSIFICATION RESULTS OF COMPETING METHODS AND THE PROPOSED BMR ON HAND. RESULTS ARE SHOWN IN THE FORM OF "MEAN(STANDARD DEVIATION)" AND THE BEST RESULTS ARE SHOWN IN BOLD.

Method	ANI vs. HC classification on HAND						ICH vs. HC classification on HAND					
	AUC (%)	ACC (%)	FI (%)	SEN (%)	SPE (%)	BAC (%)	AUC (%)	ACC (%)	FI (%)	SEN (%)	SPE (%)	BAC (%)
SVM	61.73(3.21) *	57.28(2.68)	56.25(2.98)	58.25(4.27)	57.22(2.87)	57.73(2.72)	53.91(3.54) *	52.52(4.75)	51.86(5.61)	53.95(6.29)	51.50(7.25)	52.73(5.14)
XGBoost	56.72(3.59) *	53.34(2.83)	51.11(2.82)	51.70(3.61)	56.70(6.59)	54.21(2.76)	55.05(3.66) *	53.04(2.19)	51.29(2.27)	52.34(3.46)	55.55(3.51)	53.95(1.92)
Random Forest	64.64(1.46)	58.61(1.54)	57.88(3.03)	60.49(5.13)	59.51(5.45)	60.00(1.52)	57.11(4.23) *	53.10(4.79)	51.70(3.95)	53.31(4.81)	55.89(8.04)	54.60(4.70)
MLP	64.68(3.82) *	58.97(2.69)	57.37(2.49)	59.05(2.82)	59.55(3.85)	59.30(2.28)	56.44(4.02)	55.25(3.40)	55.81(2.55)	59.30(2.09)	52.24(6.86)	55.77(3.40)
GCN	64.23(2.13) *	59.60(2.32)	57.78(3.29)	58.65(5.06)	61.65(4.89)	60.15(2.39)	55.42(6.44) *	54.65(1.49)	53.08(5.09)	55.60(9.52)	53.63(10.20)	54.62(1.60)
GIN	65.93(3.61) *	60.34(2.25)	58.40(3.06)	59.00(4.57)	62.36(3.49)	60.68(2.72)	57.78(3.60) *	54.80(3.69)	53.78(4.11)	55.90(4.58)	56.39(4.04)	56.14(3.21)
GAT	65.72(3.86) *	61.11(2.99)	59.24(0.78)	**62.53(5.83)**	59.47(6.00)	61.00(0.09)	54.42(3.01) *	53.67(4.03)	54.41(4.00)	60.06(8.78)	50.12(7.77)	55.09(4.57)
BrainGNN	63.04(3.23) *	60.26(2.60)	56.87(3.78)	56.32(5.95)	63.70(5.59)	60.01(1.78)	58.14(4.75)	56.74(2.23)	**57.38(2.84)**	**61.48(3.80)**	51.20(3.45)	56.34(2.36)
STGCN	53.81(2.94) *	51.20(4.00)	52.45(3.78)	57.40(5.45)	47.40(5.64)	52.40(4.02)	55.22(8.36) *	51.11(7.62)	50.88(10.36)	55.34(22.93)	49.48(8.16)	52.41(8.40)
BolT	61.70(8.88) *	59.92(8.51)	58.01(11.35)	58.92(16.26)	60.86(12.69)	59.89(8.59)	57.15(8.94)	54.20(6.64)	53.16(8.13)	54.18(11.51)	54.29(10.88)	54.23(6.66)
BMR (Ours)	**67.91(3.49)**	**64.03(4.97)**	**61.78(6.89)**	62.39(8.68)	**66.25(3.81)**	**64.32(5.16)**	**60.26(4.98)**	**57.49(4.15)**	55.15(5.58)	54.82(7.26)	**61.12(6.94)**	**57.97(4.34)**

'*' THE TERM REPRESENTS THE PROPOSED BMR IS STATISTICALLY SIGNIFICANTLY DIFFERENT FROM A COMPETING METHOD.

TABLE IV

PERFORMANCE OF THE PROPOSED BMR AND ITS THREE VARIANTS THAT USE DIFFERENT SIMILARITY METRICS IN THE PROPOSED MODULARITY CONSTRAINT IN ASD VS. HC CLASSIFICATION ON NYU.

Method	AUC (%)	ACC (%)	FI (%)	SEN (%)	SPE (%)	BAC (%)
BMR-ED	73.00(1.69)	65.88(2.86)	60.62(3.33)	64.37(3.80)	68.19(2.51)	66.28(2.40)
BMR-HD	72.03(2.69)	64.75(2.57)	59.04(3.21)	60.88(4.21)	68.25(4.20)	64.57(2.52)
BMR-JS	73.12(3.01)	66.92(4.54)	61.48(4.85)	63.13(4.80)	**70.53(6.20)**	66.83(4.61)
BMR	**73.24(4.48)**	**67.10(4.29)**	**62.16(4.53)**	**64.74(5.08)**	69.39(6.22)	**67.06(4.38)**

TABLE V

THE TOP 10 MOST DISCRIMINATIVE FCs, RELATED REGIONS-OF-INTEREST (ROIs), AND CORRESPONDING *p*-VALUES IN ASD VS. HC CLASSIFICATION ON NYU FROM ABIDE AND ANI VS. HC CLASSIFICATION ON HAND. NOTE THAT BRAIN ROIs ARE DIVIDED BASED ON AUTOMATED ANATOMICAL LABELING (AAL) ATLAS.

ASD vs. HC classification on NYU			**AM vs. HC classification on HAND**		
FC	**Related ROIs**	**p-value |**	**FC**	**ROI name**	***p*-value**
(32, 63)	(ACG.R, SMG.L)	5.35×10^{-5}	(19, 30)	(SMA.R, INS.R)	3.31×10^{-5}
(9, 65)	(ORBmid.L, ANG.L)	6.85×10^{-5}	(12, 72)	(IFGoperc.R, CAU.R)	3.04×10^{-4}
(38, 90)	(HIRR, ITG.R)	7.77×10^{-5}	(21, 69)	(OLF.L, PCL.L)	3.37×10^{-4}
(107, 109)	(CRBL10.L, Vermis 12)	9.55×10^{-4}	(32, 101)	(ACG.R, CRBL7b.L)	3.73×10^{-4}
(40, 111)	(PHG.R, Vermis45)	1.26×10^{-4}	(28, 66)	(REC.R, ANG.R)	7.40×10^{-4}
(48, 112)	(LING.R, Vermis6)	2.04×10^{-4}	(95, 108)	(CRBL3.L, CRBL10.R)	8.19×10^{-4}
(18, 75)	(ROL.R, PAL.L)	2.82×10^{-4}	(53, 81)	(IOG.L, STG.L)	9.39×10^{-4}
(28, 31)	(REC.R, ACG.L)	2.93×10^{-4}	(27, 83)	(REC.L, REC.L)	1.24×10^{-4}
(27, 31)	(REC.L, ACG.L)	2.99×10^{-4}	(70, 98)	(PCL.R, CRBL45.R)	1.25×10^{-4}
(28, 32)	(REC.R, ACG.R)	3.18×10^{-4}	(29, 55)	(INS.L, FFG.L)	1.57×10^{-4}