Modularity-Constrained Dynamic Representation Learning for Interpretable Brain Disorder Analysis with Functional MRI

Qianqian Wang[1], Mengqi Wu[1], Yuqi Fang[1], Wei Wang[2], Lishan Qiao[3(✉)], and Mingxia Liu[1(✉)]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
mingxia_liu@med.unc.edu

[2] Department of Radiology, Beijing Youan Hospital, Capital Medical University, Beijing 100069, China

[3] School of Mathematics Science, Liaocheng University, Shandong 252000, China
qiaolishan@lcu.edu.cn

Abstract. Resting-state functional MRI (rs-fMRI) is increasingly used to detect altered functional connectivity patterns caused by brain disorders, thereby facilitating objective quantification of brain pathology. Existing studies typically extract fMRI features using various machine/deep learning methods, but the generated imaging biomarkers are often challenging to interpret. Besides, the brain operates as a modular system with many cognitive/topological modules, where each module contains subsets of densely inter-connected regions-of-interest (ROIs) that are sparsely connected to ROIs in other modules. However, current methods cannot effectively characterize brain modularity. This paper proposes a modularity-constrained dynamic representation learning (MDRL) framework for interpretable brain disorder analysis with rs-fMRI. The MDRL consists of 3 parts: (1) dynamic graph construction, (2) modularity-constrained spatiotemporal graph neural network (MSGNN) for dynamic feature learning, and (3) prediction and biomarker detection. In particular, the MSGNN is designed to learn spatiotemporal dynamic representations of fMRI, constrained by 3 functional modules (*i.e.*, central executive network, salience network, and default mode network). To enhance discriminative ability of learned features, we encourage the MSGNN to reconstruct network topology of input graphs. Experimental results on two public and one private datasets with a total of 1,155 subjects validate that our MDRL outperforms several state-of-the-art methods in fMRI-based brain disorder analysis. The detected fMRI biomarkers have good explainability and can be potentially used to improve clinical diagnosis.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_5.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 46–56, 2023.
https://doi.org/10.1007/978-3-031-43907-0_5

Keywords: Functional MRI · Modularity · Biomarker · Brain disorder

1 Introduction

Resting-state functional magnetic resonance imaging (rs-fMRI) has been increasingly used to help us understand pathological mechanisms of neurological disorders by revealing abnormal or dysfunctional brain connectivity patterns [1–4]. Brain regions-of-interest (ROIs) or functional connectivity (FC) involved in these patterns can be used as potential biomarkers to facilitate objective quantification of brain pathology [5]. Previous studies have designed various machine and deep learning models to extract fMRI features and explore disease-related imaging biomarkers [6,7]. However, due to the complexity of brain organization and black-box property of many learning-based models, the generated biomarkers are usually difficult to interpret, thus limiting their utility in clinical practice [8,9].

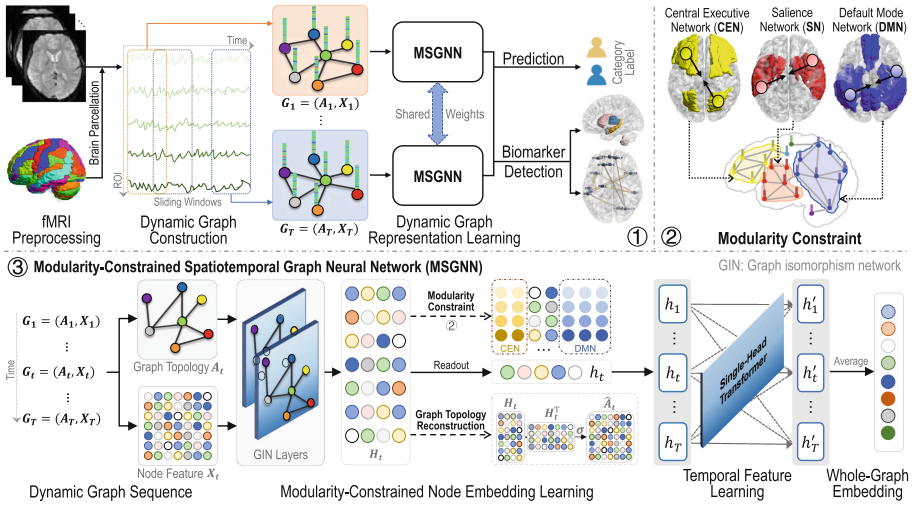

Fig. 1. Illustration of our modularity-constrained dynamic representation learning (MDRL) framework, with 3 components: (1) dynamic graph construction via sliding windows, (2) modularity-constrained spatiotemporal graph neural network (MSGNN) for dynamic representation learning, and (3) prediction and biomarker detection. The MSGNN is designed to learn spatiotemporal features via GIN and transformer layers, constrained by 3 neurocognitive modules (*i.e.*, central executive network, salience network, and default mode network) and graph topology reconstruction.

On the other hand, the human brain operates as a modular system, where each module contains a set of ROIs that are densely connected within the module but sparsely connected to ROIs in other modules [10,11]. In particular, the central executive network (CEN), salience network (SN), and default mode network (DMN) are three prominent resting-state neurocognitive modules in the

brain, supporting efficient cognition [12]. Unfortunately, existing learning-based fMRI studies usually ignore such inherent modular brain structures [13,14].

To this end, we propose a modularity-constrained dynamic representation learning (**MDRL**) framework for interpretable brain disorder analysis with rs-fMRI. As shown in Fig. 1, the MDRL consists of (1) dynamic graph construction, (2) modularity-constrained spatiotemporal graph neural network (MSGNN) for dynamic graph representation learning, and (3) prediction and biomarker detection. The MSGNN is designed to learn spatiotemporal features via graph isomorphism network and transformer layers, constrained by 3 neurocognitive modules (*i.e.*, central executive network, salience network, and default mode network). To enhance discriminative ability of learned fMRI embeddings, we also encourage the MSGNN to reconstruct topology of input graphs. To our knowledge, this is among the first attempts to incorporate modularity prior to graph learning models for fMRI-based brain disorder analysis. Experimental results on two public and one private datasets validate the effectiveness of the MDRL in detecting three brain disorders with rs-fMRI data.

2 Materials and Methodology

2.1 Subjects and Image Preprocessing

Two public datasets (*i.e.*, **ABIDE** [15] and **MDD** [16]) and one private HIV-associated neurocognitive disorder (**HAND**) dataset with rs-fMRI are used. The two largest sites (*i.e.*, NYU and UM) of ABIDE include 79 patients with autism spectrum disorder (ASD) and 105 healthy controls (HCs), and 68 ASDs and 77 HCs, respectively. The two largest sites (*i.e.*, Site 20 and Site 21) of MDD contain 282 patients with major depressive disorder (MDD) and 251 HCs, 86 MDDs and 70 HCs, respectively. The HAND were collected from Beijing YouAn Hospital, with 67 asymptomatic neurocognitive impairment patients (ANIs) with HIV and 70 HCs. Demographics of subjects are reported in *Supplementary Materials*.

All rs-fMRI data were preprocessed using the Data Processing Assistant for Resting-State fMRI (DPARSF) pipeline [17]. Major steps include (1) magnetization equilibrium by trimming the first 10 volumes, (2) slice timing correction and head motion correction, (3) regression of nuisance covariates (*e.g.*, white matter signals, ventricle, and head motion parameters), (4) spatial normalization to the MNI space, (5) bandpass filtering (0.01–0.10 Hz). The average rs-fMRI time series of 116 ROIs defined by the AAL atlas are extracted for each subject.

2.2 Proposed Method

As shown in Fig. 1, the proposed MDRL consists of (1) dynamic graph construction via sliding windows, (2) MSGNN for dynamic graph representation learning, and (3) prediction and biomarker detection, with details introduced below.

Dynamic Graph Construction. Considering that brain functional connectivity (FC) patterns change dynamically over time [18], we propose to first construct dynamic networks/graphs using a sliding window strategy for each subject. Denote original fMRI time series as $S \in R^{N \times M}$, where N is the number of ROIs and M is the number of time points of blood-oxygen-level-dependent (BOLD) signals in rs-fMRI. We first divide the original 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. Then, we construct an FC network by calculating Pearson correlation (PC) coefficients between time series of pairwise ROIs for each of T segments, denoted as $X_t \in R^{N \times N} (t = 1, \cdots, T)$. The original feature for the j-th node is represented by the j-th row in X_t for segment t. Considering all connections in an FC network may include some noisy or redundant information, we retain the top 30% strongest edges in each FC network to generate an adjacent matrix $A_t \in \{0, 1\}^{N \times N}$ for segment t. Thus, the obtained dynamic graph sequence of each subject can be described as $G_t = \{A_t, X_t\}$ $(t = 1, \cdots, T)$.

Modularity-Constrained Spatiotemporal GNN. With the dynamic graph sequence $\{G_t\}_{t=1}^{T}$ as input, we design a modularity-constrained spatiotemporal graph neural network (MSGNN) to learn interpretable and discriminative graph embeddings, with two unique constraints: 1) a *modularity constraint*, and 2) a *graph topology reconstruction constraint*. In MSGNN, we first stack two graph isomorphism network (GIN) layers [19] for node-level feature learning. The node-level embedding H_t at the segment t learned by GIN layers is formulated as:

$$H_t = \psi(\varepsilon^{(1)} I + A_t)[\psi(\varepsilon^{(0)} I + A_t) X_t W^{(0)}] W^{(1)} \tag{1}$$

where ψ is nonlinear activation, $\varepsilon^{(i)}$ is a parameter at the i-th GIN layer, I is an identity matrix, and $W^{(i)}$ is the weight for the fully connected layers in GIN.

1) Modularity Constraint. It has been demonstrated that the central executive network (CEN), salience network (SN) and default mode network (DMN) are three crucial neurocognitive modules in the brain and these three modules have been consistently observed across different individuals and experimental paradigms, where CEN performs high-level cognitive tasks (*e.g.*, decision-making and rule-based problem-solving), SN mainly detects external stimuli and coordinates brain neural resources, and DMN is responsible for self-related cognitive functions [10–12]. The ROIs/nodes within a module are densely inter-connected, resulting in a high degree of clustering between nodes from the same module. Based on such prior knowledge and clinical experience, we reasonably assume that the *learned embeddings of nodes within the same neurocognitive module tend to be similar*. We develop a novel modularity constraint to encourage similarity between paired node-level embeddings in the same module. Mathematically, the proposed *modularity constraint* is formulated as:

$$L_M = -\sum\nolimits_{t=1}^{T} \sum\nolimits_{k=1}^{K} \sum\nolimits_{i,j=1}^{N_k} \frac{h_i^{t,k} \cdot h_j^{t,k}}{\|h_i^{t,k}\| \cdot \|h_j^{t,k}\|} \tag{2}$$

where $h_i^{t,k}$ and $h_j^{t,k}$ are embeddings of two nodes in the k-th module (with N_k ROIs) at segment t, and K is the number of modules ($K = 3$ in this work). With Eq. (2), we encourage the MSGNN to focus on modular brain structures during representation learning, thus improving discriminative ability of fMRI features.

2) Graph Topology Reconstruction Constraint. To further enhance discriminative ability of learned embeddings, we propose to preserve graph topology by reconstructing adjacent matrices. For the segment t, its adjacent matrix A_t can be reconstructed through $\hat{A}_t = \sigma(H_t \cdot H_t^\top)$, where σ is a nonlinear mapping function. The *graph topology reconstruction constraint* is then formulated as:

$$L_R = \sum\nolimits_{t=1}^{T} \Psi(A_t, \hat{A}_t) \tag{3}$$

where Ψ is a cross-entropy loss function. We then apply an SERO operation [20] to generate graph-level embeddings based on node-level embeddings, formulated as $h_t = H_t \Phi(P^{(2)} \sigma(P^{(1)} H_t \phi_{mean}))$, where Φ is a sigmoid function, $P^{(1)}$ and $P^{(2)}$ are learnable weight matrices, and ϕ_{mean} is average operation.

3) Temporal Feature Learning. To further capture temporal information, a single-head transformer is used to fuse features derived from T segments, with a self-attention mechanism to model temporal dynamics across segments. We then sum the learned features $\{h_i^{'}\}_{i=1}^{T}$ to obtain the whole-graph embedding.

Prediction and Biomarker Detection. The whole-graph embedding is fed into a fully connected layer with Softmax for prediction, with final loss defined as:

$$L = L_C + \lambda_1 L_R + \lambda_2 L_M \tag{4}$$

where L_C is a cross-entropy loss for prediction, and λ_1 and λ_2 are two hyper-parameters. To facilitate interpretation of our learned graph embeddings, we calculate PC coefficients between paired node embeddings for each segment and average them across segments to obtain an FC network for each subject. The upper triangle of each FC network is flattened into a vector and Lasso [21] (with default parameter) is used to select discriminative features. Finally, we map these features to the original feature space to detect disease-related ROIs and FCs.

Implementation. The MDRL is implemented in PyTorch and trained using an Adam optimizer (with learning rate of 0.001, training epochs of 30, batch size of 8 and $\tau = 20$). We set window size $\Gamma = 40$ for NYU and $\Gamma = 70$ for the rest, and results of MDRL with different Γ values are shown in *Supplementary Materials*. In the modularity constraint, we randomly select $m = 50\%$ of all $\frac{N_k(N_k-1)}{2}$ paired ROIs in the k-th module (with N_k ROIs) to constrain the MDRL.

3 Experiment

Competing Methods. We compare the MDRL with 2 shallow methods: 1) linear **SVM** with node-level statistics (*i.e.*, degree centrality, clustering coefficient, betweenness centrality, and eigenvector centrality) of FC networks as fMRI features (with each FC network constructed using PC), 2) **XGBoost** with the same features as SVM; and 4 state-of-the-art (SOTA) deep models with default architectures: 3) **GCN** [22], 4) **GAT** [23], 5) **BrainGNN** [9], and 6) **STGCN** [18].

Table 1. Results of seven methods on ABIDE.

Method	ASD vs. HC classification on NYU					ASD vs. HC classification on UM				
	AUC (%)	ACC (%)	SEN (%)	SPE (%)	BAC (%)	AUC (%)	ACC (%)	SEN (%)	SPE (%)	BAC (%)
SVM	56.6(2.9)*	54.8(3.1)	51.5(4.6)	57.9(4.7)	54.7(3.1)	53.6(4.3)*	53.3(3.6)	50.3(4.5)	56.6(4.7)	53.5(3.8)
XGBoost	61.9(0.6)*	63.0(1.6)	48.0(2.7)	**75.9(3.7)**	61.9(0.6)	58.8(0.8)*	58.6(1.9)	47.6(3.6)	70.0(3.9)	58.8(0.8)
GCN	67.5(3.3)*	63.6(3.1)	51.0(5.1)	73.5(4.7)	62.3(2.8)	66.7(2.6)	60.0(3.0)	54.6(4.4)	66.5(5.2)	60.6(2.6)
GAT	64.9(2.6)*	60.1(2.6)	53.0(4.9)	66.1(3.2)	59.5(2.8)	66.5(3.5)*	60.4(3.1)	**56.1(3.1)**	65.4(6.9)	60.7(2.8)
BrainGNN	66.9(2.9)*	63.2(3.2)	**57.1(4.8)**	68.5(3.0)	62.8(3.2)	65.9(2.5)	62.7(2.6)	55.5(3.3)	68.1(5.7)	61.8(2.1)
STGCN	66.6(0.8)*	61.5(1.5)	53.6(2.3)	68.4(1.7)	61.0(1.5)	64.0(0.1)*	63.9(0.1)	55.9(1.1)	72.1(0.5)	64.0(0.1)
MDRL (Ours)	**72.6(1.7)**	**65.6(2.1)**	57.0(2.8)	74.1(3.1)	**65.6(1.9)**	**67.1(2.3)**	**64.5(1.4)**	55.6(3.9)	**72.7(2.5)**	**64.1(2.4)**

The term '*' denotes the results of MDRL and a competing method are statistically significantly different ($p < 0.05$).

Table 2. Results of seven methods on MDD.

Method	MDD vs. HC classification on Site 20					MDD vs. HC classification on Site 21				
	AUC (%)	ACC (%)	SEN (%)	SPE (%)	BAC (%)	AUC (%)	ACC (%)	SEN (%)	SPE (%)	BAC (%)
SVM	53.7(2.1)	53.0(2.6)	54.6(3.1)	51.5(1.8)	53.0(2.4)	52.8(3.0)*	52.7(2.7)	59.0(3.8)	45.8(4.7)	52.4(2.8)
XGBoost	55.9(1.9)*	55.9(2.2)	**64.4(4.1)**	47.4(1.3)	55.9(1.9)	52.0(2.3)*	52.5(2.7)	**66.2(4.7)**	37.8(4.6)	52.0(2.3)
GCN	55.7(2.7)	54.9(2.1)	59.6(4.4)	50.1(4.4)	54.8(2.0)	54.8(3.1)*	54.0(3.0)	60.9(6.0)	46.6(7.5)	53.8(3.3)
GAT	57.8(1.3)*	55.7(1.3)	61.4(5.7)	49.5(4.4)	55.4(1.1)	53.2(3.1)*	52.8(2.4)	62.0(4.7)	42.7(7.1)	52.3(3.4)
BrainGNN	56.3(2.4)*	52.8(2.1)	51.7(7.6)	**55.0(8.5)**	53.4(2.0)	53.9(3.3)	53.5(8.6)	58.4(12.8)	45.7(2.0)	52.1(5.4)
STGCN	54.2(4.5)*	54.6(4.1)	56.6(5.2)	52.2(2.9)	54.4(4.0)	54.9(0.3)*	53.4(0.1)	61.4(1.8)	44.2(3.7)	52.7(0.9)
MDRL (Ours)	**60.9(2.6)**	**57.4(1.9)**	62.2(5.0)	51.6(3.3)	**56.9(1.4)**	**56.6(9.4)**	**55.2(7.8)**	58.1(10.2)	**51.3(7.1)**	**54.6(7.8)**

The term '*' denotes the results of MDRL and a competing method are statistically significantly different ($p < 0.05$).

Experimental Setting. Three classification tasks are performed: 1) ASD vs. HC on ABIDE, 2) MDD vs. HC on MDD, and 3) ANI vs. HC on HAND. A 5-fold cross-validation (CV) strategy is employed. Within each fold, we also perform an inner 5-fold CV to select optimal parameters. Five evaluation metrics are used: area under ROC curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), and balanced accuracy (BAC). Paired sample t-test is performed to evaluate whether the MDRL is significantly different from a competing method.

Classification Results. Results achieved by different methods in three classification tasks on three datasets are reported in Tables 1–2 and Fig. 2. It can be seen that our MDRL generally outperforms two shallow methods (*i.e.*, SVM and XGBoost) that rely on handcrafted node features without modeling whole-graph topological information. Compared with 4 SOTA deep learning methods, our MDRL achieves superior performance in terms of most metrics in three tasks. For instance, for ASD vs. HC classification on NYU of ABIDE (see Table 1), the

AUC value of MDRL is improved by 5.7% compared with BrainGNN (a SOTA method designed for brain network analysis). This implies the MDRL can learn discriminative graph representations to boost fMRI-based learning performance.

Ablation Study. We compare the proposed MDRL with its three degenerated variants: 1) **MDRLw/oM** without the modularity constraint, 2) **MDRLw/oR** without the graph topology reconstruction constraint, and 3) **MDRLw/oMR** without the two constraints. The results are reported in Fig. 3, from which one can see that MDRL is superior to its three variants, verifying the effectiveness of the two constraints defined in Eqs. (2)–(3). Besides, MDRLw/oM is generally inferior to MDRLw/oR in three tasks, implying that the modularity constraint may contribute more to MDRL than the graph reconstruction constraint.

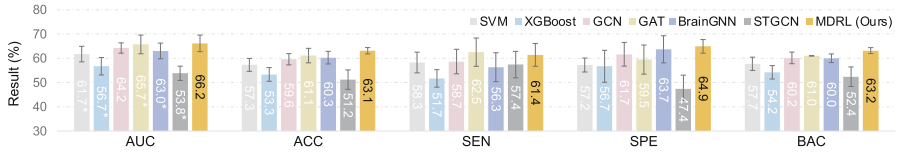

Fig. 2. Results of seven methods in ANI vs. HC classification on HAND.

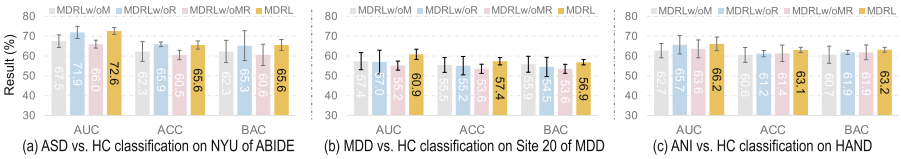

Fig. 3. Performance of the MDRL and its variants in three tasks on three datasets.

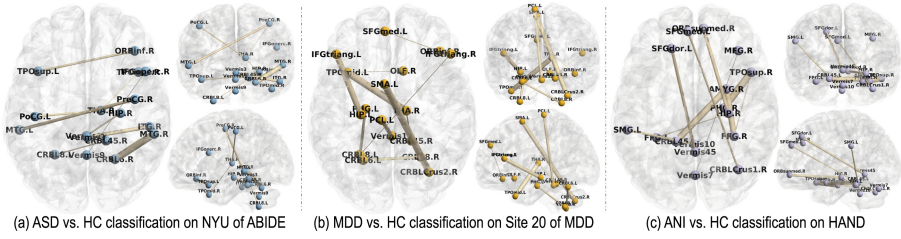

Fig. 4. Visualization of the top 10 most discriminative functional connectivities identified by the MDRL in identifying 3 diseases on 3 datasets (with AAL for ROI partition).

Discriminative ROIs and Functional Connectivities. The top 10 discriminative FCs detected by the MDRL in three tasks are shown in Fig. 4. The thickness of each line represents discriminative power that is proportional to the corresponding Lasso coefficient. For ASD identification (see Fig. 4 (a)), the FCs involved in *thalamus* and *middle temporal gyrus* are frequently identified, which complies with previous findings [24,25]. For MDD detection (see Fig. 4 (b)), we find that several ROIs (*e.g.*, *hippocampus*, *supplementary motor area* and *thalamus*) are highly associated with MDD identification, which coincides with previous studies [26–28]. For ANI identification (see Fig. 4 (c)), the detected ROIs such as *amygdala*, *right temporal pole: superior temporal gyrus* and *parahippocampal gyrus*, are also reported in previous research [29–31]. This further demonstrates the effectiveness of the MDRL in disease-associated biomarker detection.

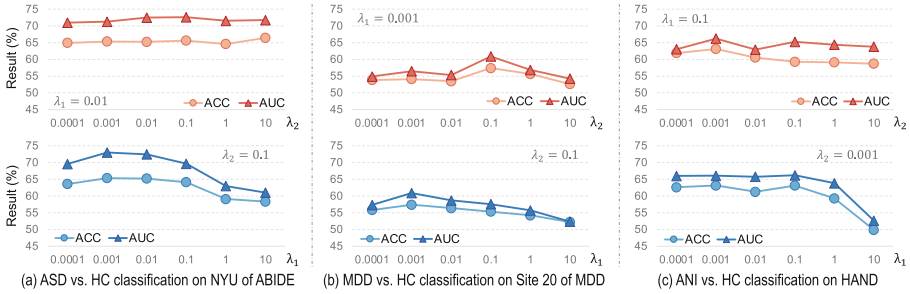

Fig. 5. Results of our MDRL with different hyperparameters (*i.e.*, λ_1 and λ_2).

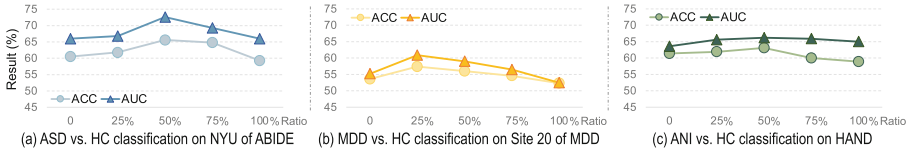

Fig. 6. Results of the proposed MDRL with different modularity ratios.

4 Discussion

Parameter Analysis. To investigate the influence of hyperparameters, we vary the values of two parameters (*i.e.*, λ_1 and λ_2) in Eq. (4) and report the results of MDRL in Fig. 5. It can be seen from Fig. 5 that, with λ_1 fixed, the performance of MDRL exhibits small fluctuations with the increase of parameter values of λ_2, implying that MDRL is not very sensitive to λ_2 in three tasks. With λ_2 fixed, the MDRL with a large λ_1 (*e.g.*, $\lambda_1 = 10$) achieves worse performance. The possible reason could be that using a strong graph reconstruct constraint will make the model difficult to converge, thus degrading its learning performance.

Influence of Modularity Ratio. In the main experiments, we randomly select $m = 50\%$ of all $\frac{N_k(N_k-1)}{2}$ paired ROIs in the k-th module (with N_k ROIs) to constrain the MDRL. We now vary the modularity ratio m within $[0\%, 25\%, \cdots, 100\%]$ and record the results of MDRL in three tasks in Fig. 6. It can be seen from Fig. 6 that, when $m < 75\%$, the ACC and AUC results generally increase as m increases. But when using a large modularity ratio (*e.g.*, $m = 100\%$), the MDRL cannot achieve satisfactory results. This may be due to the over-smoothing problem caused by using a too-strong modularity constraint.

Influence of Network Construction. We use PC to construct the original FC networks in MDRL. We also use sparse representation (SR) and low-rank representation (LR) for network construction in MDRL and report results in Table 3. It can be seen from Table 3 that the MDRL with PC outperforms its two variants. The underlying reason could be that PC can model dependencies among regional BOLD signals without discarding any connection information.

Table 3. Results of the MDRL with different FC network construction strategies.

Method	ASD vs. HC on NYU of ABIDE			MDD vs. HC on Site 20 of MDD			ANI vs. HC on HAND		
	AUC (%)	ACC (%)	BAC (%)	AUC (%)	ACC (%)	BAC (%)	AUC (%)	ACC (%)	BAC (%)
MDRL_LR	62.2(3.5)	59.2(4.6)	59.4(4.6)	54.4(5.0)	53.5(4.9)	53.2(4.8)	58.2(3.3)	60.6(2.0)	60.7(1.8)
MDRL_SR	60.9(1.9)	62.7(1.9)	60.8(2.7)	55.5(7.4)	52.7(3.8)	53.2(4.6)	64.6(4.0)	61.0(2.2)	61.2(2.0)
MDRL	**72.6(1.7)**	**65.6(2.1)**	**65.6(2.0)**	**60.9(2.6)**	**57.4(1.9)**	**56.9(1.4)**	**66.2(3.4)**	**63.1(1.3)**	**63.2(1.3)**

5 Conclusion and Future Work

In this work, we propose a modularity-constrained dynamic graph representation (MDRL) framework for fMRI-based brain disorder analysis. We first construct dynamic graphs for each subject and then design a modularity-constrained GNN to learn spatiotemporal representation, followed by prediction and biomarker detection. Experimental results on three rs-fMRI datasets validate the superiority of the MDRL in brain disease detection. Currently, we only characterize pairwise relationships of ROIs within 3 prominent neurocognitive modules (*i.e.*, CEN, SN, and DMN) as prior knowledge to design the modularity constraint in MDRL. Fine-grained modular structure and disease-specific modularity constraint will be considered. Besides, we will employ advanced harmonization methods [32] to reduce inter-site variance, fully utilizing multi-site fMRI data for model training.

Acknowledgment. M. Wu, Y. Fang, and M. Liu were supported by an NIH grant RF1AG073297.

References

1. Pagani, M., et al.: mTOR-related synaptic pathology causes autism spectrum disorder-associated functional hyperconnectivity. Nat. Commun. **12**(1), 6084 (2021)

2. Sezer, I., Pizzagalli, D.A., Sacchet, M.D.: Resting-state fMRI functional connectivity and mindfulness in clinical and non-clinical contexts: a review and synthesis. Neurosci. Biobehav. Rev. (2022) 104583
3. Liu, J., et al.: Astrocyte dysfunction drives abnormal resting-state functional connectivity in depression. Sci. Adv. **8**(46), eabo2098 (2022)
4. Sahoo, D., Satterthwaite, T.D., Davatzikos, C.: Hierarchical extraction of functional connectivity components in human brain using resting-state fMRI. IEEE Trans. Med. Imaging **40**(3), 940–950 (2020)
5. Traut, N., et al.: Insights from an autism imaging biomarker challenge: promises and threats to biomarker discovery. Neuroimage **255**, 119171 (2022)
6. Azevedo, T., et al.: A deep graph neural network architecture for modelling spatio-temporal dynamics in resting-state functional MRI data. Med. Image Anal. **79**, 102471 (2022)
7. Bessadok, A., Mahjoub, M.A., Rekik, I.: Graph neural networks in network neuroscience. IEEE Trans. Pattern Anal. Mach. Intell. (2022)
8. Zhang, Z., Xie, Y., Xing, F., McGough, M., Yang, L.: MDNet: a semantically and visually interpretable medical image diagnosis network. In: CVPR, pp. 6428–6436 (2017)
9. Li, X., et al.: BrainGNN: interpretable brain graph neural network for fMRI analysis. Med. Image Anal. **74**, 102233 (2021)
10. Sporns, O., Betzel, R.F.: Modular brain networks. Annu. Rev. Psychol. **67**, 613–640 (2016)
11. Bertolero, M.A., Yeo, B.T., D'Esposito, M.: The modular and integrative functional architecture of the human brain. Proc. Natl. Acad. Sci. **112**(49), E6798–E6807 (2015)
12. Goulden, N., et al.: The salience network is responsible for switching between the default mode network and the central executive network: replication from DCM. Neuroimage **99**, 180–190 (2014)
13. Geirhos, R., et al.: Shortcut learning in deep neural networks. Nat. Mach. Intell. **2**(11), 665–673 (2020)
14. Knyazev, B., Taylor, G.W., Amer, M.: Understanding attention and generalization in graph neural networks. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
15. Di Martino, A., et al.: The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol. Psychiatry **19**(6), 659–667 (2014)
16. Yan, C.G., et al.: Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. Proc. Natl. Acad. Sci. **116**(18), 9078–9083 (2019)
17. Yan, C., Zang, Y.: DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. Front. Syst. Neurosci. **4**, 13 (2010)
18. Gadgil, S., Zhao, Q., Pfefferbaum, A., Sullivan, E.V., Adeli, E., Pohl, K.M.: Spatio-temporal graph convolution for resting-state fMRI analysis. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 528–538. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_52
19. Kim, B.H., Ye, J.C.: Understanding graph isomorphism network for rs-fMRI functional connectivity analysis. Front. Neurosci. 630 (2020)
20. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: CVPR, pp. 7132–7141 (2018)
21. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. Roy. Stat. Soc.: Ser. B (Methodol.) **58**(1), 267–288 (1996)

22. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
23. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
24. Ayub, R., et al.: Thalamocortical connectivity is associated with autism symptoms in high-functioning adults with autism and typically developing adults. Transl. Psychiatry **11**(1), 93 (2021)
25. Xu, J., et al.: Specific functional connectivity patterns of middle temporal gyrus subregions in children and adults with autism spectrum disorder. Autism Res. **13**(3), 410–422 (2020)
26. MacQueen, G., Frodl, T.: The hippocampus in major depression: evidence for the convergence of the bench and bedside in psychiatric research? Mol. Psychiatry **16**(3), 252–264 (2011)
27. Sarkheil, P., Odysseos, P., Bee, I., Zvyagintsev, M., Neuner, I., Mathiak, K.: Functional connectivity of supplementary motor area during finger-tapping in major depression. Compr. Psychiatry **99**, 152166 (2020)
28. Batail, J.M., Coloigner, J., Soulas, M., Robert, G., Barillot, C., Drapier, D.: Structural abnormalities associated with poor outcome of a major depressive episode: the role of thalamus. Psychiatry Res. Neuroimaging **305**, 111158 (2020)
29. Clark, U.S., et al.: Effects of HIV and early life stress on amygdala morphometry and neurocognitive function. J. Int. Neuropsychol. Soc. **18**(4), 657–668 (2012)
30. Zhan, Y., et al.: The resting state central auditory network: a potential marker of HIV-related central nervous system alterations. Ear Hear. **43**(4), 1222 (2022)
31. Sarma, M.K., et al.: Regional brain gray and white matter changes in perinatally HIV-infected adolescents. NeuroImage Clin. **4**, 29–34 (2014)
32. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Trans. Biomed. Eng. **69**(3), 1173–1185 (2021)