Multi-modal Dynamic Graph Network: Coupling Structural and Functional Connectome for Disease Diagnosis and Classification

Yanwu Yang[1, 2], Xutao Guo[1, 2], Zhikai Chang[1], Chenfei Ye[2, 4], Yang Xiang[2*], and Ting Ma[1, 2, 3, 4*]

[1]Electronic & Informatin Engineering School, Harbin Institute of Technology (Shenzhen), China
[2]Peng Cheng Laboratory, Shenzhen, China
[3]Guangdong Provincial Key Laboratory of Aerospace Communication and Networking Technology,
Harbin Institute of Technology (Shenzhen), Shenzhen, China
[4]International Research Institute for Artifcial Intelligence,
Harbin Institute of Technology (Shenzhen), Shenzhen, China

Abstract—**Multi-modal neuroimaging technology has greatly facilitated the diagnosis efficiency a nd d iagnosis a ccuracy, and provides complementary information in discovering objective disease biomarkers. Conventional deep learning methods, e.g. convolutional neural networks, overlook relationships between nodes and fail to capture topological properties in graphs. Graph neural networks have been proven to be of great importance in modeling brain connectome networks and relating disease-specific p atterns. H owever, m ost e xisting g raph m ethods explicitly require known graph structures, which are not available in the sophisticated brain system. Especially in heterogeneous multi-modal brain networks, there exists a great challenge to model interactions among brain regions in consideration of inter-modal dependencies. In this study, we propose a Multi-modal Dynamic Graph Convolution Network (MDGCN) for structural and functional brain network learning. Our method benefits f rom m odeling i nter-modal r epresentations a nd relating attentive multi-model associations into dynamic graphs with a compositional correspondence matrix. Moreover, a bilateral graph convolution layer is proposed to aggregate multi-modal representations in terms of multi-modal associations. Extensive experiments on three datasets demonstrate the superiority of our proposed method in terms of disease classification, with the accuracy of 90.4%, 85.9% and 98.3% in predicting Mild Cognitive Impairment, Parkinson's Disease, and Schizophrenia respectively. Our statistical evaluations on the correspondence matrix exhibit a high correspondence with previous evidence of biomarkers.**

Index Terms—**Graph Neural Network, Multi-modal Graph Network, Diagnosis, Dynamic Graph**

I. INTRODUCTION

Recently, computer-aided diagnosis technologies using advanced neuroimaging developments have been widely adopted for medical scenarios, e.g. disease diagnosis, and medical image segmentation. Functional Magnetic Resonance Imaging (fMRI) and Diffusion Tensor Imaging (DTI) have become promising candidates for brain study. Functional MRI is a stimulus-free acquisition used to track changes in co-activation across brain regions. DTI captures the directional diffusion of water molecules as a proxy for structural connectivity. Derived functional and structural connectivity is feasible to model the brain as a network by representing brain parcellations along with their structural or functional connectivity. The brain connectome provides a more holistic view by modeling the entire human brain and characterizes individual subject behavior, cognition, and mental health [1]. There is mounting evidence demonstrates that functional and structural connectivity could identify predictive biomarkers for brain disorders such as Alzheimer's disease (AD), Schizophrenia (SCZ), and Parkinson's disease (PD) [2]–[4].

Medical image-based diagnosis is a challenging task due to the sophisticated structure of brain systems and subtle lesions, which might be overlooked by medical experts [5]. Neuroimage processing with multiple modalities is feasible to assess and develop distinctive biomarkers from multiple fields. Previous studies link the functional signals with structural pathways for mediating and suggest that functional connectivity and structural connectivity might be mediated by each other [6], [7]. Recently, state-of-the-art graph neural networks (GNN) have achieved promising performance in multi-modal graph-structural data learning [7], [8]. However, most existing GNNs built graphs on the originally derived connectivity, and fail to sufficiently model sophisticated associations among nodes. This issue is even aggravated when modeling multi-modal brain networks since there exist heterogeneous structures and representations among multiple modalities. However, most existing studies potentially ignore these issues and achieves sub-optimal results.

In terms of this, we propose a Multi-modal Dynamic Graph Convolution Network (MDGCN) to model multi-modal complementary associations by dynamic graphs. Our network allows for tighter coupling of context between multiple modalities by representing functional and structural connectome dynamically and providing a compositional space for reasoning. Specially, we first parse both the functional and structural into dynamic graphs with embedded representations as nodes. A correspondence factor matrix is introduced to

* Corresponding authors.

capture the corresponding values of each pair of nodes between modalities, which is denoted as the adjacency matrix. And multimodal representations are aggregated by a Bilateral Graph Convolution (BGC) layer for complementary message passing. Extensive experiments on three datasets demonstrate that our proposed method outperforms other baselines in the prediction of Mild Cognitive Impairment (MCI), Parkinson's Disease (PD), and Schizophrenia (SCHZ) with the accuracy performances of 90.4%, 85.9%, and 98.3% respectively.

The rest of our paper is structured as follows. We would like to review competitive methods in terms of connectome study and multi-modal models in Section II. The details of the proposed model are introduced in Section III. Section IV describes the experiments of our proposed model in disease classification on 3 datasets and provides the experimental results. Section V draws the conclusions of the work.

II. RELATED WORKS

A. Brain connectome network study

With the flexibility of uncovering the complex biological mechanisms using rs-fMRI and DTI, deep learning methods have been widely coordinated to examine and analyze the patterns. Convolution neural networks (CNN) and graph neural networks (GNN) have become useful tools for brain connectome embedding, where high dimensional neuroimaging features are embedded into a low dimensional space that preserves their context as well as capturing topological attributes. BrainNetCNN is proposed to take the brain connectome networks as grid-like data, and measure the topological locality in connectome [9], which has achieved promising performance for disease diagnosis and phenotype prediction.

Apart from convolution neural networks, graph neural networks retain a state that can represent information about the neighbors and provides a powerful way to explore the dependencies between nodes. However, applying a graph network directly to the brain connectome is problematic. On one hand, brain networks have sophisticated and non-linear structures. For example, most existing methods apply the derived functional connectivity as the adjacency matrix, which is measured linearly between two brain regions. These derived linear connectivity fails to model complex associations between brain regions. On the other hand, graph convolution networks explicitly require a known graph structure, which is not available in the brain connectome. Several strategies have been proposed to tackle the unknown structure issue [10], [11]. Especially, dynamic graph convolution methods are proposed to model graph structures adaptively to characterize intrinsic brain connectome representations and achieve promising performances in prediction [12], [13]. Nevertheless, there is still a lack of studies to tackle the multi-modal connectome graphs.

B. Multi-modal connectome learning

Existing multi-modal connectome learning methods can be categorized into two classes: feature learning methods and deep learning methods. Compared with feature learning methods that leverage feature selection to identify disease-related features, deep learning methods are feasible to capture intrinsic meaningful representations and achieves better performances. [14] devised a calibration mechanism to fuse fMRI and DTI information into edges. [15] proposed to perform a two-layer convolution on the fMRI and DTI data simultaneously. [7] regularizes convolution on functional connectivity with structural graph Laplacian. However, most of these studies lack the ability to sufficiently model complementary associations between modalities, since there is a lack of joint compositional reasoning over both functional and structural connectome networks.

III. METHOD

The proposed Multi-modal Dynamic Graph Convolution Network (MDGCN) aims at parsing multi-modal representations into dynamic graphs and performing graph aggregation for message passing. In this section, we would like to firstly introduce the brain graph definition, and then the detail of the proposed method.

A. Preliminaries

Brain network graph: The brain networks derived from neuroimages are usually symmetric positive define (SPD) matrices $X \in \mathbb{R}^{M \times M}$, where M denotes the number of brain regions. Each element $x_{i,j}$ denotes a co-variance or connectivity strength between the regions. The brain network is usually formulated as an undirected graph $G = (V, E, H)$, where V is a finite set of vertices with $|V| = M$ and $E \in \mathbb{R}^{M \times M}$ denotes the edges in the graphs. The nodes and edges are represented by the derived SPD matrices X. For each vertex v_i, the node feature vector h_i is constructed by the i-th row or column in the SPD matrix $h_i = \{x_{i,k} | k = 1, 2, ..., M\}$. The edges are represented by the matrices directly, of which an element is assigned by $e_{i,j} = x_{i,j}$.

Multi-modal brain graph: The multi-modal brain graphs are constructed by the functional and structural brain networks derived from fMRI and DTI respectively. An input $\hat{G}$ is expressed by a tuple of graphs as $\hat{G} = \{G^s, G^f\}$, where G^s and G^f denote the structural and functional brain network graphs respectively. Formally, given a set of graphs $\{\hat{G}_1, \hat{G}_2, ..., \hat{G}_N\}$ with a few labeled graph instances, the aim of the study is to decide the state of the unlabeled graphs as a graph classification task.

Dynamic graph: Since there remain unknown node relationships in the multi-modal graphs, applying the graph convolution on the multi-modal graphs is problematic. The dynamic graphs leverage the dynamic mechanism to model multi-modal node representations and interactions by learning mappings: $f_V : X \rightarrow V, f_E : X \rightarrow E$. And the learned graphs are denoted as dynamic graphs.

B. Dynamic Multi-modal Graph Mapping

The key to brain graph mapping is how to parse multi-modal graph representations. In this study, we propose to model the node representations as a sequence via the Gated Recurrent Unit (GRU), where the gating mechanism allows
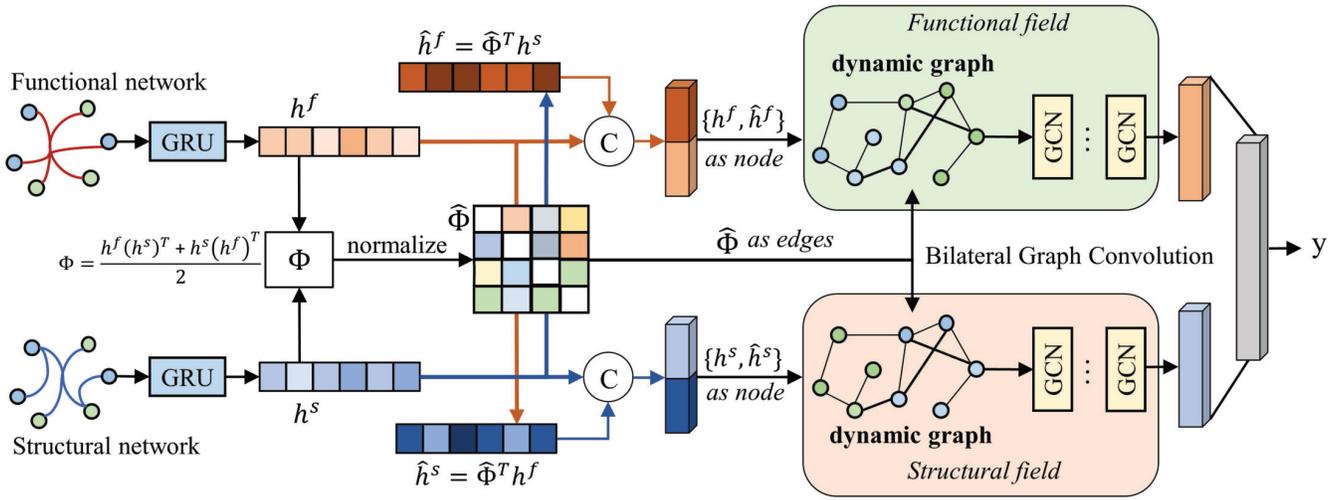
Fig. 1. Illustration of the proposed MDGCN method, where GRU denotes a gated recurrent unit for node representation learning, GCN denotes a graph convolution layer. The multi-modal representations are cross embedded and fed into a bilateral graph convolution module. Message passing on the graphs are guided by the correspondence matrix.

for the learning of sequential relationships and protect the learning from undesired updates [16]. The GRU layers take the input brain network matrix $X \in \mathbb{R}^{M \times M}$ as a sequence with M nodes and M features. Formally, given an input $X = \{X^f, X^s\}$, where X^f and X^s denotes the functional and structural brain networks, the node features are embedded by:

$$h_j^s = GRU(\hat{x}^s), h_j^f = GRU(\hat{x}^f) \tag{1}$$

Specially, given an input $\hat{x}$, the GRU operation on each parcel $k \in [1, M]$ is formulated as:

$$z_k = \sigma(W_z \cdot [o_{k-1}, \hat{x}_k]) \tag{2}$$

$$r_k = \sigma(W_r \cdot [o_{k-1}, \hat{x}_k]) \tag{3}$$

$$\hat{h}_k = tanh(W \cdot [r_k * o_{k-1}, \hat{x}_k]) \tag{4}$$

$$o_k = (1 - z_k) * o_{k-1} + z_k * \hat{o}_k \tag{5}$$

where, $\hat{x}_k$ denotes the input vector corresponding to the k-th brain region, and z_k, r_k are the update gate and reset gate respectively. With the embedded h_j^f and h_j^s, a soft correspondence matrix Φ is obtained by:

$$\Phi = \frac{h_j^f (h_j^s)^T + h_j^s (h_j^f)^T}{2} \tag{6}$$

Each row vector in Φ is a probability distribution over potential correspondences to corresponding nodes. The matrix can be regarded as the scores for measuring the goodness of matches between nodes in two modalities. A sinkhorn function is applied to normalize the matrix, which satisfies doubly stochastic, where $\sum_j^M \hat{\Phi}_{i,j} = 1$.

By obtaining the normalized correspondence matrix $\hat{\Phi}$, we can project the representations from one source into another source by:

$$\hat{h}^f = \hat{\Phi}^T h^s \tag{7}$$

$$\hat{h}^s = \hat{\Phi}^T h^f \tag{8}$$

With the obtained embedded representations $\hat{h}^f$ and $\hat{h}^s$, the dynamic node features can be built by $\{\hat{h}^f, h^f\}$ or $\{\hat{h}^s, h^s\}$, which represents the transformation representations of functional/structural field. Moreover, we formulate the normalized correspondence matrix $\hat{\Phi}$ as the dynamic adjacency matrix.

C. Bilateral Graph Convolution

In this study, a Bilateral Graph Convolution (BGC) is proposed to perform convolution on the multi-modal graphs. To tackle the heterogeneous features between modalities, the BGC module applies convolutions on each field to aggregate representations of each single modality separately. Moreover, we implement the spatial aggregation on the graphs for message passing instead of spectral graph convolution. Since the brain network is fully connected, graph spatial convolution as well as spectral graph convolutions are able to aggregate global information. In this way, the graph spatial convolution is formulated with $\hat{\Phi}$ as:

$$H_{l+1}^f = \sigma(\hat{\Phi} H_l^f W_l^f), \text{where } H_0^f = ||\{\hat{h}^f, h^f\} \tag{9}$$

$$H_{l+1}^s = \sigma(\hat{\Phi} H_l^s W_l^s), \text{where } H_0^s = ||\{\hat{h}^s, h^s\} \tag{10}$$

$||$ denotes a concatenation operation, σ denotes a sigmoid activation function, and W is a learnable matrix for improving node representations. The outputs of the BGC layer are further combined as a feature vector, and fed into a multi-layer perception for classification.

D. Optimization

The final output is further fed into a three-layer multi-layer perception classifier followed with a ReLU activation function and a dropout layer. A softmax function is implemented for probability output. The loss function is formulated as the cross-entropy function. To summarize, the detail of the process of our proposed DMGN is shown in Algorithm 1:

Algorithm 1: Dynamic Multi-modal Graph Network

Input: Multi-modal brain networks
$\{\hat{G}_1, \hat{G}_2, \hat{G}_3, ..., \hat{G}_n\}$, where $\hat{G} = \{G^s, G^f\}$;
For each graph $G = (V, E, X)$

Output: Prediction y of the test set

1 Calculate dynamic node features by GRU:
   $h^s = GRU(\hat{x}^s)$, $h^f = GRU(\hat{x}^f)$;

2 Calculate the correspondence matrix:
   $\Phi = \frac{1}{2}h_j^f(h_j^s)^T + \frac{1}{2}h_j^s(h_j^f)^T$;

3 Normalize Φ into $\hat{\Phi}$ to satisfy doubly stochastic;

4 Obtain the cross-modality mapping representations:

5 $\hat{h}^f = \hat{\Phi}^T h^s$;

6 $\hat{h}^s = \hat{\Phi}^T h^f$;

7 **for** l=1:L **do**

8    $H_l^f = \sigma(\hat{\Phi}H_{l-1}^f W_{l-1}^f)$, where $H_0^f = ||\{\hat{h}^f, h^f\}$;

9    $H_l^s = \sigma(\hat{\Phi}H_{l-1}^s W_{l-1}^s)$, where $H_0^s = ||\{\hat{h}^s, h^s\}$;

10 **end**

11 Readout $H_L = ||\{H_L^f, H_L^s\}$;

12 $y = softmax(H_L)$

IV. EXPERIMENTS AND RESULTS

A. Datasets

In this study, three real-world datasets are employed in this study, where functional MRI and DTI are aggregated. All the datasets are enrolled for multi-modal graph classification.

ADNI Dataset[1]: The ADNI dataset is a longitudinal multimodal neuroimaging dataset. In this study, we collected 114 subjects that were diagnosed at the baseline for evaluation including 51 healthy controls (NC) and 63 mild cognitive impairment (MCI). Notably, MCI is considered to be a significant stage for preclinical diagnosis of AD.

Xuanwu dataset [2]: A total of 155 subjects are included in this dataset, where 70 HCs and 85 subjects with Parkinson's Disease (PD) were recruited from the Movement Disorders Clinic of the Xuanwu Hospital of Capital Medical University. The patients were diagnosed according to the MDS Clinical Diagnostic Criteria for Parkinson's disease.

CHUV dataset [17]: The MRI data were obtained from the Service of General Psychiatry at the Lausanne University Hospital including 27 healthy participants and 27 schizophrenic patients. All of the patients were diagnosed with schizophrenic disorders after meeting the DSM-IV criteria.

---

[1]http://www.adni-info.org/

B. Preprocessing

All the fMRI images were pre-processed by reference to the Configurable Pipeline for the Analysis of Connectomes (CPAC) pipeline [18], including skull striping, slice timing correction, motion correction, global mean intensity normalization, nuisance signal regression with 24 motion parameters, and band-pass filtering (0.01-0.08Hz). The functional images were finally registered into a standard anatomical space (MNI-152). The mean time series for a set of regions were computed and normalized into zero mean and unit variance. The Pearson Coefficient Correlation is applied to measure the functional connectivity.

The DTI images were pre-processed by image denoising, head motion, eddy-current, susceptibility distortion, and field inhomogeneity correction by MRtrix 3 [19]. The fiber orientation distributions were estimated by constrained spherical deconvolution. We performed the 2-nd order Integration over Fiber Orientation Distributions to reconstruct 10 million streamlines. A Spherical-deconvolution Informed Filtering of Tractograms was applied to reduce the streamline count to 5 million. The number of streamlines connecting each pair of brain regions was used to construct the structural network.

All the pre-processed fMRI and DTI images were mapped by the brain template for parcellations. In this study, the images in ADNI and Xuanwu datasets were segmented by the Schaefer atlas, which was parceled by a gradient weighted Markov random field approach that identified 100 cortical parcels. CHUV data were parcellated into 83 cortical and subcortical areas by the Freesurfer.

C. Implementation details

In our implementation, the number of layers of graph convolution is decided in a grid search from 1 to 4. The output of graph convolution is further fed into a 3-layer multilayer perception classifier followed by a leaky ReLU activation function and a dropout layer. The learning rate is set as 3e-4, and the weight decay is 5e-5. All the models in this study are trained for 600 epochs and would be early stopped then the loss has not been decreased for 100 epochs. We trained the models with PyTorch on two NVIDIA 2080-Ti GPUs. For better comparison, 10-fold cross-validation was applied for evaluation by randomly sampling 90% data for training and 10% for testing in each fold. For all experiments, we evaluated the performance in terms of the diagnosis accuracy (Acc), precision (Prec), and Area Under the Curve (AUC).

D. Competitive methods

In this study, we compare our proposed MDGCN with well-estimated graph methods and models that are designed specifically for brain connectome. These methods include:

SVM and MLP. The conventional machine learning methods of the support vector machine and multi-layer perception are compared as a baseline of classification. The upper matrix of the brain networks are fed into the classifiers to give a score for each subject. The layer number of MLP is searched from 1 to 4.

| Method	ADNI			Xuanwu			CHUV		
	Acc	Prec	AUC	Acc	Prec	AUC	Acc	Prec	AUC
SVM	63.2±9.5	60.8±6.5	68.3±13.6	62.6±3.3	90.6±2.5	69.8±5.1	64.7±4.2	63.4±7.0	72.9±15.9
M-MLP	80.6±9.9	76.29±12.3	83.9±9.8	73.8±7.4	70.2±6.4	73.6±9.7	82.7±11.1	80.0±18.9	82.4±11.6
BrainNetCNN	82.3±11.5	82.1±14.1	82.8±9.5	75.6±9.7	71.4±12.1	77.6±10.7	92.7±9.0	95.0±10.0	94.7±8.7
M-GCN	83.3±7.4	80.9±15.9	81.7±13.6	81.3±4.6	77.1±11.3	75.4±10.2	90.7±12.9	94.2±11.8	86.8±19.7
HGNN	81.1±6.5	84.8±14.2	85.6±6.4	77.7±9.8	76.9±16.2	80.3±12.2	96.0±8.0	100.0±0.0	98.9±0.9
DHGNN	84.4±6.0	84.2±12.3	86.7±5.4	78.2±6.8	80.1±13.1	80.8±8.1	90.7±9.4	83.3±17.9	92.4±10.8
DMGN(Ours)	90.4±2.4	91.6±6.9	88.2±6.6	85.9±4.5	85.5±6.0	83.1±8.1	98.3±5.0	100.0±0.0	99.2±1.7
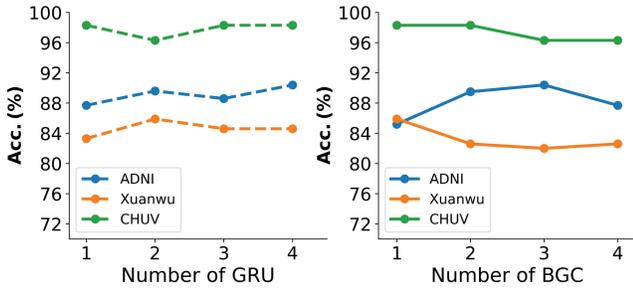
Fig. 2. The effect of the number setting for the GRU layer and Bilateral Graph Convolution layer.


Fig. 3. Comparison results on the node embedding and graph convolution.

BrainNetCNN [9]. BrainNetCNN is a CNN-based framework for brain network study with promising performances in brain network study. The model is implemented by multiple convolution layers to learn the multi-modal inputs.

M-GCN [7]. M-GCN is a multi-modal graph convolution network that aggregates functional representations by regularizing with structural graph Laplacian, which outperforms several stat-of-the-art baselines in predicting phenotype values.

HGNN [20]. A hypergraph graph neural network is implemented to encode hyper structure. The multi-modal brain networks can be fed into the HGNN as a nature of hyperedges.

DHGNN [13]. The dynamic graph hyperneural network extends the HGNN into a dynamic graph. These two methods are implemented to compare with the hyperedge mechanism for modeling multi-modal data.

E. Parameter setting

The settings of the GRU layer and the bilateral graph convolution layer are discussed in Figure 2, where the results on the ADNI, Xuanwu and CHUV datasets are shown in blue, yellow, and green respectively. The prediction performances would slightly change with the setting on the layer number of GRU. Moreover, as the layer number of bilateral graph convolution increases, the performance trends to increase firstly and then decrease. We suspect that this is caused by the global aggregation of the graph convolution, which leads to over-smoothing on graphs.

F. Multi-modal graph classification performance

Table I demonstrates the comparison results, where the best results are shown in bold, and the second best are in underline. It is shown that conventional machine learning SVM achieves the worst performance. This indicates that the multi-modal graphs might have nonlinear and heterogeneous structures, and a simply linear classifier fails to distinguish the brain disease state. Moreover, compared with SVM, multi-layer perception improves the performances significantly, where nonlinear representations are deeply embedded with multiple layers. BrainNetCNN and M-GCN achieve further improvements, showing that deep learning methods such as CNN and GNN are feasible to capture and learn more meaningful representations. In addition, the hypergraph neural network HGNN performs worse than BrainNetCNN and M-GCN on the ADNI and Xuanwu datasets. While DHGNN outperforms these baseline models on the ADNI and Xuanwu datasets, where the dynamic mechanism might contribute to richer representations. Moreover, our proposed MDGCN achieves the best performances among all the competitive models in terms of the accuracy, precision and AUC, with 6.0%, 4.6% and 2.3% improvements on the ADNI, Xuanwu and CHUV datasets respectively. We suspect that the improvements are caused by the powerful cross-modality correspondence reasoning that captures inter-modal dependencies to aggregate multi-modal representations.
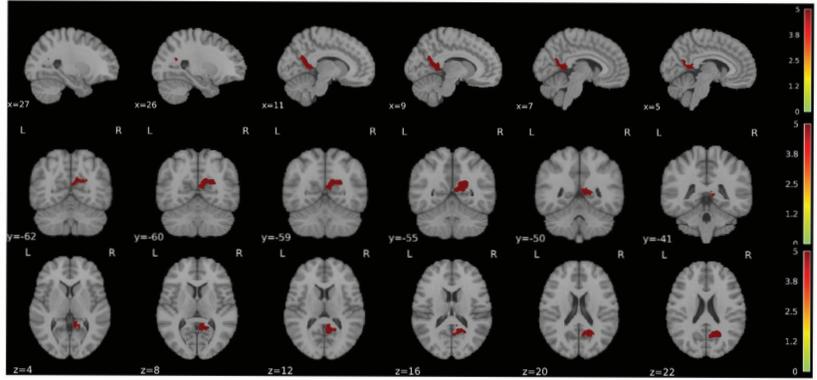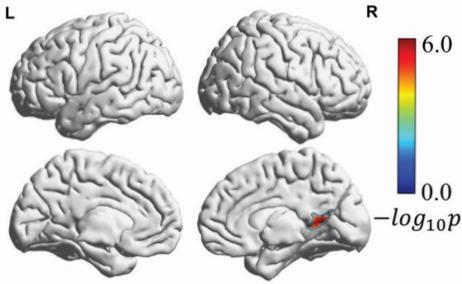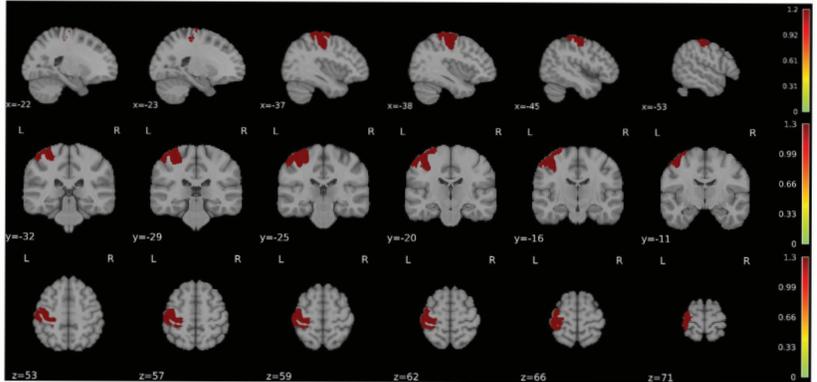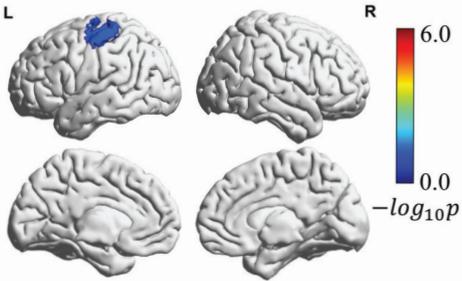
Fig. 4. Statistical results ($-log_{10}p$) on the correspondence matrix, where only significant regions are displayed. The right posterior cingulate gyrus in the default mode network, the left somatomotor network are located in MCI and PD respectively.

G. Ablation studies

Ablation studies were carried out to measure the importance of the components in the MDGCN, including edge-aware convolution and the correspondence matrix. Figure 3 plots the comparison results on the way of node embedding and the correspondence matrix. On one hand, the GRU for node embedding is compared with other node embedding methods, including multi-layer perception, graph embedding, and LSTM. As is shown in Figure 3 A), the LSTM outperforms the graph embedding and MLP methods. Moreover, the GRU performs better than LSTM on the ADNI and Xuanwu datasets. The results demonstrate that the gating mechanism in LSTM and GRU could relate long-term attentive relationships, capture more meaningful node representations, and protect the node representations from undesired updates, which plays a key role in dynamic graph node representation learning. The graph embedding and MLP embed all the node features in the same and fails to consider the importance of connectivity patterns. On the other hand, we also evaluated the effect of the correspondence matrix by comparing with the single domain aggregation. In detail, the equation (6) is modified as $\Phi = h_j^f(h_j^s)^T$ and $\Phi = h_j^s(h_j^f)^T$ to calculate the function-structure and structure-function transformation respectively. And only a single domain aggregation is read out by $H_L = H_L^f$ or $H_L = H_L^s$. Figure 3 B) displays the comparison results. It

is shown that results of graph aggregations in the functional domain and structural domain are comparable in terms of the accuracy. Moreover, our method incorporating cross-domain aggregation outperforms the single domain aggregation. This indicates that the interplay between modalities might be a key in modeling multi-modal representations.

H. Biological explanation

The correspondence matrices model the multi-modal interactions among regions, which are feasible to be applied to pinpoint the key brain biomarkers for disease. In this section, we extracted the correspondence matrix for each subject and performed a group statistical analysis for the explanation. The experiments were carried out on the ADNI and Xuanwu datasets that consist of more than 100 subjects for example, which could tell the power of the proposed DMGCN in interpretation ability. In detail, we performed the state-of-the-art statistical method, multi-distance multi-variant regression (MDMR), for group comparison on the correspondence matrix. The Bonferroni correction was applied to control the false positive rate. And the p value < 0.05 after correction was determined significant within the experiments.

Figure 4 demonstrates the statistical analysis of the two datasets, where the regions with significant differences (p value < 0.05) are displayed with the value of $-log_{10}p$. The

regions of the right posterior cingulate gyrus in the default mode network and the left somatomotor were found in the ADNI, and Xuanwu datasets respectively. Previous studies demonstrate that the AD process has been hypothetically explained by PCC hypofunction, due to the effect of the degeneration of cingulum fibers [21]. And PCC hypofunction could be caused by early PCC atrophy, which has been proven to be a useful biomarker [22]. Moreover, Parkinson's Disease is shown to be beginning to develop somatomotor dysfunctions with deficits in neocortical association areas [23]. Our findings in the right posterior cingulate gyrus of the default mode network, and the left somatomotor network coincide with these studies, indicating that our proposed MDGCN could locate meaningful and interpretative biomarkers.

V. CONCLUSION

In this study, we propose a multi-modal dynamic graph convolution network for disease diagnosis and classification, which exploits a dynamic graph for tighter coupling of multi-modal representations. The derived correspondence matrix provides a compositional space for reasoning multi-modal dependencies. The experimental results on three datasets demonstrate that our proposed method is feasible to model multi-modal graphs and outperforms other state-of-the-art methods. Moreover, by performing statistical analysis on the correspondence matrix, the high correspondence with previous evidence of biomarkers exhibits the interpretation ability of our proposed MDGCN, which provides a powerful way of multi-modal brain network learning.

ACKNOWLEDGMENT

REFERENCES

[1] K. Dadi, M. Rahim, A. Abraham, D. Chyzhyk, M. Milham, B. Thirion, G. Varoquaux, A. D. N. Initiative *et al.*, "Benchmarking functional connectome-based predictive models for resting-state fmri," *NeuroImage*, vol. 192, pp. 115–134, 2019.

[2] Y. Yang, C. Ye, J. Sun, L. Liang, H. Lv, L. Gao, J. Fang, T. Ma, and T. Wu, "Alteration of brain structural connectivity in progression of parkinson's disease: a connectome-wide network analysis," *NeuroImage: Clinical*, vol. 31, p. 102715, 2021.

[3] C. Ye, S. Mori, P. Chan, and T. Ma, "Connectome-wide network analysis of white matter connectivity in alzheimer's disease," *NeuroImage: Clinical*, vol. 22, p. 101690, 2019.

[4] E. Damaraju, E. A. Allen, A. Belger, J. M. Ford, S. McEwen, D. Mathalon, B. Mueller, G. Pearlson, S. Potkin, A. Preda *et al.*, "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *NeuroImage: Clinical*, vol. 5, pp. 298–308, 2014.

[5] E. Yang, M. Liu, D. Yao, B. Cao, C. Lian, P.-T. Yap, and D. Shen, "Deep bayesian hashing with center prior for multi-modal neuroimage retrieval," *IEEE transactions on medical imaging*, vol. 40, no. 2, pp. 503–513, 2020.

[6] M. Fukushima, R. F. Betzel, Y. He, M. P. van den Heuvel, X.-N. Zuo, and O. Sporns, "Structure–function relationships during segregated and integrated network states of human brain functional connectivity," *Brain Structure and Function*, vol. 223, no. 3, pp. 1091–1106, 2018.

[7] N. S. Dsouza, M. B. Nebel, D. Crocetti, J. Robinson, S. Mostofsky, and A. Venkataraman, "M-gcn: A multimodal graph convolutional network to integrate functional and structural connectomics data to predict multidimensional phenotypic characterizations," in *Medical Imaging with Deep Learning*. PMLR, 2021, pp. 119–130.

[8] R. Saqur and K. Narasimhan, "Multimodal graph networks for compositional generalization in visual question answering," *Advances in Neural Information Processing Systems*, vol. 33, pp. 3070–3081, 2020.

[9] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, "Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, 2017.

[10] H. Hu, D. Ji, W. Gan, S. Bai, W. Wu, and J. Yan, "Class-wise dynamic graph convolution for semantic segmentation," in *European Conference on Computer Vision*. Springer, 2020, pp. 1–17.

[11] Y. Zhang and H. Huang, "New graph-blind convolutional network for brain connectome data analysis," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 669–681.

[12] K. Zhao, B. Duka, H. Xie, D. J. Oathes, V. Calhoun, and Y. Zhang, "A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in adhd," *Neuroimage*, vol. 246, p. 118774, 2022.

[13] J. Jiang, Y. Wei, Y. Feng, J. Cao, and Y. Gao, "Dynamic hypergraph neural networks." in *IJCAI*, 2019, pp. 2635–2641.

[14] X. Song, F. Zhou, A. F. Frangi, J. Cao, X. Xiao, Y. Lei, T. Wang, and B. Lei, "Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction," *Medical Image Analysis*, vol. 69, p. 101947, 2021.

[15] Y. Wang, Y. Yang, X. Guo, C. Ye, N. Gao, Y. Fang, and H. T. Ma, "A novel multimodal mri analysis for alzheimer's disease based on convolutional neural network," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 754–757.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[17] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium *et al.*, "The wu-minn human connectome project: an overview," *Neuroimage*, vol. 80, pp. 62–79, 2013.

[18] C. Craddock, S. Sikka, B. Cheung, R. Khanuja, S. S. Ghosh, C. Yan, Q. Li, D. Lurie, J. Vogelstein, R. Burns *et al.*, "Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac)," *Front Neuroinform*, vol. 42, pp. 10–3389, 2013.

[19] J.-D. Tournier, F. Calamante, and A. Connelly, "Mrtrix: diffusion tractography in crossing fiber regions," *International journal of imaging systems and technology*, vol. 22, no. 1, pp. 53–66, 2012.

[20] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 3558–3565.

[21] I. H. Choo, D. Y. Lee, J. S. Oh, J. S. Lee, D. S. Lee, I. C. Song, J. C. Youn, S. G. Kim, K. W. Kim, J. H. Jhoo *et al.*, "Posterior cingulate cortex atrophy and regional cingulum disruption in mild cognitive impairment and alzheimer's disease," *Neurobiology of aging*, vol. 31, no. 5, pp. 772–779, 2010.

[22] T. C. Chua, W. Wen, X. Chen, N. Kochan, M. J. Slavin, J. N. Trollor, H. Brodaty, and P. S. Sachdev, "Diffusion tensor imaging of the posterior cingulate is a useful biomarker of mild cognitive impairment," *The American Journal of Geriatric Psychiatry*, vol. 17, no. 7, pp. 602–613, 2009.

[23] H. Braak, J. R. Bohl, C. M. Müller, U. Rüb, R. A. de Vos, and K. Del Tredici, "Stanley fahn lecture 2005: The staging procedure for the inclusion body pathology associated with sporadic parkinson's disease reconsidered," *Movement disorders: official journal of the Movement Disorder Society*, vol. 21, no. 12, pp. 2042–2051, 2006.