

Multi-Scale FC-Based Multi-Order GCN: A Novel Model for Predicting Individual Behavior From fMRI

Xuyun Wen¹, Qumei Cao, Bin Jing², and Daoqiang Zhang¹, *Senior Member, IEEE*

Abstract—Predicting individual behavior from brain imaging data using machine learning is a rapidly growing field in neuroscience. Functional connectivity (FC), which captures interactions between different brain regions, contains valuable information about the organization of the brain and is considered a crucial feature for modeling human behavior. Graph convolutional networks (GCN) have proven to be a powerful tool for extracting graph structure features and have shown promising results in various FC-based classification tasks, such as disease classification and prognosis prediction. Despite this success, few behavior prediction models currently exist based on GCN, and their performance is not satisfactory. To address this gap, a new model called the Multi-Scale FC-based Multi-Order GCN (MSFC-MO-GCN) was proposed in this paper. The model considers the hierarchical structure of the brain system and utilizes FCs inferred from multiple spatial scales as input to comprehensively characterize individual brain organization. To enhance the feature learning ability of GCN, a multi-order graph convolutional layer is incorporated, which uses multi-order neighbors to guide message passing and learns high-order graph information of nodal connections. Additionally, an inter-subject contrast constraint is designed to control the potential information redundancy of FCs among different spatial scales during the feature learning process. Experimental evaluation were conducted on the publicly available dataset from human connectome project. A total of 805 healthy subjects were included and 5 representative behavior metrics were used. The experimental results show that our proposed method outperforms the existing behavior prediction models in all behavior prediction tasks.

Index Terms—Functional connectivity, human behavior, graph convolutional network, multi-scale, multi-order.

I. INTRODUCTION

A FUNDAMENTAL goal of neuroscience is to understand how the structure and function of the human brain

Manuscript received 27 July 2023; revised 26 December 2023; accepted 17 January 2024. Date of publication 22 January 2024; date of current version 29 January 2024. This work was supported in part by the China Post-Doctoral Science Foundation under Grant 2021TQ0150 and Grant 2021M701699; in part by the National Natural Science Foundation of China under Grant 62001222 and Grant 62136004; and in part by the 2021 Jiangsu Shuangchuang Talent Program. (*Corresponding authors: Xuyun Wen; Daoqiang Zhang.*)

Xuyun Wen, Qumei Cao, and Daoqiang Zhang are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: wenyuyun@nuaa.edu.cn; cao_qumei@nuaa.edu.cn; dqzhang@nuaa.edu.cn).

Bin Jing is with the School of Biomedical Engineering, Capital Medical University, Beijing 100069, China (e-mail: jingbin@sina.com).

Digital Object Identifier 10.1109/TNSRE.2024.3357059

give rise to behavior [1], [2], [3], [4]. Previous studies have established that individual differences in diverse brain characteristics, such as morphological features [5], structural [6], [7] and functional connectivities [8], [9], [10] etc., are associated with behavioral and demographic traits. With the increasing availability of advanced neuroimaging technology and powerful computational tools, there is growing interest in using machine learning algorithms to construct individual-level behavioral predictive models based on neuroimaging data [11], [12], [13], [14], [15], [16], [17]. These prediction models not only help to explore the relationship between subtle differences in brain characteristics and specific behaviors [11], [12], [13], [14], but also identify potential individual behavioral problems [15], [16], [17], such as cognitive or psychological disorders, enabling the earlier intervention and treatment.

Compared to structural magnetic resonance imaging (MRI), functional MRI (fMRI) directly measures the blood oxygen level-dependent (BOLD) signal of neurons [18], [19], which not only allows for recording brain activity levels over different time periods, but also characterizes brain response patterns under different external stimuli. Therefore, this technique is more suitable for revealing the neural mechanisms behind complex human behaviors [20], [21], [22]. These studies typically model the brain as a complex network system, using brain regions as nodes and functional connectivity (FC) between regions as edges, known as the FC network (FCN). Mounting evidence indicates that the rich information on regional connections and brain organization patterns contained in functional connectivity networks (FCN) can significantly enhance our understanding of various behaviors and cognitions [11], [12], [23], [24], [25]. The utility of FCN in behavior prediction is particularly notable. Firstly, FCN provides a comprehensive view of brain activity, not limited to the single brain regions. This aids in clarifying how the brain network collaboratively operates to process specific behaviors [26]. Secondly, FCN highlights the inter-relationship between brain regions, which is crucial for understanding complex cognitive functions [27], [28]. Many studies also repeatedly observed significant associations between FC patterns and behavior/cognition [11], [12], [24], [25], [29], and FC has also been shown to be an effective feature for predicting individual characteristics, including language memory ability [11], [30], attention ability [31], [32], and fluid intelligence [24], [25], [29], [33] etc. Therefore, in recent years, increasing studies have begun to use FC as feature data for human behavior research

and have constructed diverse FC based individual behavior prediction models [12], [24], [25], [29], [32], [33], [34].

Currently, the most commonly used individual behavior prediction models are implemented by combining regression methods (such as linear regression [12], [35], kernel ridge regression [11], [14], and support vector regression [13], [36]) with feature selection. Although these models have achieved good performance in predicting several important behaviors, they have two significant drawbacks. First, these models heavily rely on expert knowledge for constructing the feature database limiting their generalization ability across different tasks. For example, some models directly use FC as features [12], [13], while others use connectivity profiles extracted from FCN as features [35]. The diversity of models makes it difficult to achieve satisfactory results in predicting all behaviors. Second, almost all of these methods belong to linear models, which are based on the assumption that there is a linear relationship between brain imaging measurements and behavioral scores, and thus cannot capture the complex relationship between the brain and behavior. In recent years, deep learning methods have been successfully applied in various graph-related tasks due to their powerful feature extraction ability and complex relationship analysis ability, providing new insight for constructing FC based behavior prediction models [11], [37], [38].

Due to non-lattice-like graph structure, conventional deep neural networks (DNNs) cannot extract effective feature representations of FCNs [11], [37], [38]. Graph convolutional network (GCN) is considered as a promising solution because it can learn features via message passing of nodal features based on graph structure, preserving the topological information of FCN [11], [38], [39], [40], [41], [42]. However, the existing FC-based GCN models cannot achieve the satisfactory prediction of brain behaviors, largely hindered by the following two limitations.

First, the model only learns brain connectivity representation based on the single-scale (i.e., single spatial resolution) FCN. An increasing amount of studies indicate that the brain is a hierarchical system, and its network organization is formed by neural coordination that spans overlapping spatial scales [43], [44], [45], which is crucial for efficient information processing to support responses for different behaviors. Different brain-behavior relationships may exist at different scales, and each scale provides complementary information about various processes [45]. Therefore, the current brain representation that relies on a single network scale is almost certainly incomplete and cannot fully characterize the multi-scale and hierarchical organization of the brain system in terms of its structural and functional dynamics, which in turn affects the performance of subsequent behavior prediction models.

Second, the graph convolution layer only utilizes the latent information from 1-order neighbors of nodes, ignoring the rich information generated by far-distance functional interactions among brain regions [11], [39]. Different-order node connections describe network structures from different levels of scope, which provide us valuable information with different granularities [46]. Therefore, relying solely on low-order node

similarity or even any specific-order node similarity may not necessarily perform best on all networks and target applications. For example, in classification tasks with coarse-grained classes, higher-order approximations are likely to be more helpful than lower-order approximations [46], [47].

To tackle above two problems, this paper proposes a Multi-Scale FC based Multi-Order GCN (MSFC-MO-GCN) model for individual behavior prediction. This model uses functional interactions estimated from multiple spatial resolutions as input, and designs multi-order graph convolution layer and inter-scale contrastive constraint to learn the comprehensive representation of individual's brain connectivity. The usage of multi-scale FC provides the characterization of hierarchical brain organization. Multi-order graph convolution layer allows model updating nodal features based on multi-order neighbors instead of 1-order neighbor to introduce high-order graph information to guide message passing. Inter-scale contrastive constraint ensures the within-subject across-scale similarity of graph features during feature learning process to remove the redundant information of FCs among different spatial scales. After that, multi-scale features are fused with weight connection method and used for the final individual behavior estimation.

To sum up, our main contributions are summarized as follows:

- 1) We confirm that functional interactions from different spatial scales provide complementary information for brain-behavior relationship. Based on this conclusion, we construct individual behavior prediction model based on multiple FCNs from coarse-to-fine scales, which improves the model's prediction ability.
- 2) Our method designs multi-order graph convolution layer in GCN for feature representation learning, enabling the introduction of rich graph information from different-order node connections.
- 3) Comprehensive experiments on real-world dataset verified the superior performance of our proposed method than the other conventional methods.

The rest of the paper is organized as follows. We describe the data information and imaging preprocessing in Section II. Section III investigates the relationship between FC and human behaviors from different spatial scales. In Section IV, we provide the detailed introduction of our proposed method MSFC-MO-GCN. Section V describes the model implementation details. Section VI gives results and discussion. Finally, we conclude the whole paper in Section VII.

II. MATERIALS

A. Data Information

This study was carried out using a dataset from the publicly available Human Connectome Project (HCP) database S1200 release. It includes 1200 healthy subjects with full 3T imaging scans and behavioral tests. All participants were free of current psychiatric or neurologic illness. All rs-fMRI data were collected with eye open and relaxed fixation on a projected bright cross-hair on a dark background. Each subject has two

resting-state fMRI (rs-fMRI) sessions with different readout directions. The acquisition parameters for rs-fMRI were: TR = 720 ms, TE = 33.1 ms, flip angle = 52° , voxel size = 2 mm^3 (isotropic), 72 slices, and total volumes = 1,200 (15 min). For extensive descriptions of the imaging information, please refer to [48]. After quality control, 805 subjects were finally used in this paper.

The HCP dataset includes a battery of behavioral/cognitive tests. In this paper, we selected five representative behavior tests for evaluating the performance of the individual behavior prediction model, including one motor-related test (Endurance), one executive-function-related test (Cognitive Flexibility), one memory-related test (Episodic Memory), one language-related test (Story Difficulty Level), and a comprehensive cognitive test (Fluid Intelligence). All the tests are estimated using the NIH Cognition Battery toolbox, and the raw scores for each test are further transformed into age-adjusted scores with a mean of 100 and a standard deviation of 15 using the NIH National Norms toolbox. For the detailed description of behavior test, please refer to [49].

B. Imaging Preprocessing

All rs-fMRI data were first preprocessed by “HCP fMRI volume” minimal preprocessing pipeline [50], including the following procedures: 1) gradient distortion correction, 2) head motion correction, 3) EPI distortion correction, 4) registration to the Montreal Neurological Institute (MNI) space, 5) intensity normalization to a global mean, and 6) masking out non-brain voxels. After that, we further adopted independent component analysis (ICA) based FIX Xnoiseifier to remove artifacts from fMRI data [51], [52]. During this cleanup, 24 head motion parameters (including 6 rigid-body motion parameters, their backward temporal derivatives, and squares of those 12 time series) and “bad components” estimated from ICA were regressed from blood oxygen level dependent (BOLD) signals of each scan.

III. MULTI-SCALE BRAIN-BEHAVIOR RELATIONSHIP

As mentioned in the introduction, the motivation for utilizing multi-scale FC for individual behavior prediction is based on the previous research finding that distinct brain-behavior relationships may be present at different scales, with each scale potentially providing complementary information regarding multifaceted processes. To test this hypothesis, we designed a pre-experiment in this subsection to compare brain-behavior relationships between different spatial scales and to testify whether they exhibit scale-relevant differences. The implementation of this experiment includes two steps described as follows.

Based on the preprocessed rs-fMRI data, we first used a multiscale brain parcellation atlas to estimate functional interactions among brain regions at different spatial scales. In this paper, the multiscale brain parcellation atlas was provided by [43], which was generated by clustering analysis based on the FC patterns of all gray matter voxels of 1489 participants, according to their local and global similarity. By adjusting the resolution parameter in the clustering analysis, multiple

parcellation atlases with 100 to 1000 ROIs were generated. The spatial relationships between ROIs in these multiscale atlases can be considered as biologically meaningful brain hierarchies. This study selected representative atlases with 100, 500, and 1000 ROIs, respectively, to construct FCNs at three spatial scales of coarse, medium, and fine. At each scale, the BOLD signals of all voxels belonging to each ROI were averaged to obtain ROI signals, which were then pairwise correlated using Pearson correlation to generate the corresponding FCN. For each FCN, we set all negative correlations as 0 and only kept 5% strongest edges [11].

After obtaining the FCN for each scale, we further assigned each brain region to its corresponding functional subsystem, and calculated the system-level correlation between FC and behavior at different spatial scales. We used a system-level analysis rather than node level in this experiment for the following two reasons: 1) Traditional neuroscience studies often investigate the mechanisms of cognitive function processing in the human brain via neural circuits, and thus brain-behavior analysis based on functional systems is more interpretable; 2) A matrix at the system level, containing fewer nodes, facilitates clearer and more visually intuitive comparative results across different spatial scales. To do this, we first referred to [53] to divide the brain regions into seven functional subsystems, including the visual network (VIS), somatosensory-motor network (SM), attention network (ATT), salience network (SAL), limbic system (LIM), frontoparietal network (FP), and default mode network (DMN). For each subsystem, we divided its contained nodal brain regions into two systems according to the left and right brain, respectively. Then, for each system, we calculated the Pearson correlation between all included FC connections and behavior, and calculated the percentage of connections with significant results ($p < 0.05$ after FDR correction) out of all connections to measure the correlation score (CS) between the subsystem and behavior. For connections between different subsystems, we used the same approach to calculate the correlation between inter-system FC and behavior. Therefore, for each scale of FCN and each behavior, we finally obtained a 14×14 CS matrix, where the diagonal of the matrix represents the CSs between intra-system and behavior, and the off-diagonal values represent the CSs between inter-system FC and behavior. A higher CS value represents a higher correlation between brain connectivity and the corresponding human behavior metric.

IV. MULTI-SCALE FC BASED MULTI-ORDER GRAPH CONVOLUTIONAL NETWORK

The flowchart of the proposed individual behavior prediction model is shown in Fig. 1, which consists of four steps: (Step A) multi-scale functional connectivity estimation, (Step B) multi-order graph convolution network, (Step C) adaptive feature fusion, and (Step D) behavior score estimation. In Step A, the input is voxel-wise BOLD signal, and we construct FCNs from multiple spatial scales by utilizing a set of coarse-to-fine brain parcellations on rs-fMRI data. Each scale provides a characterization of brain connectivity from a specific spatial perspective. Step B uses multi-scale FCNs as

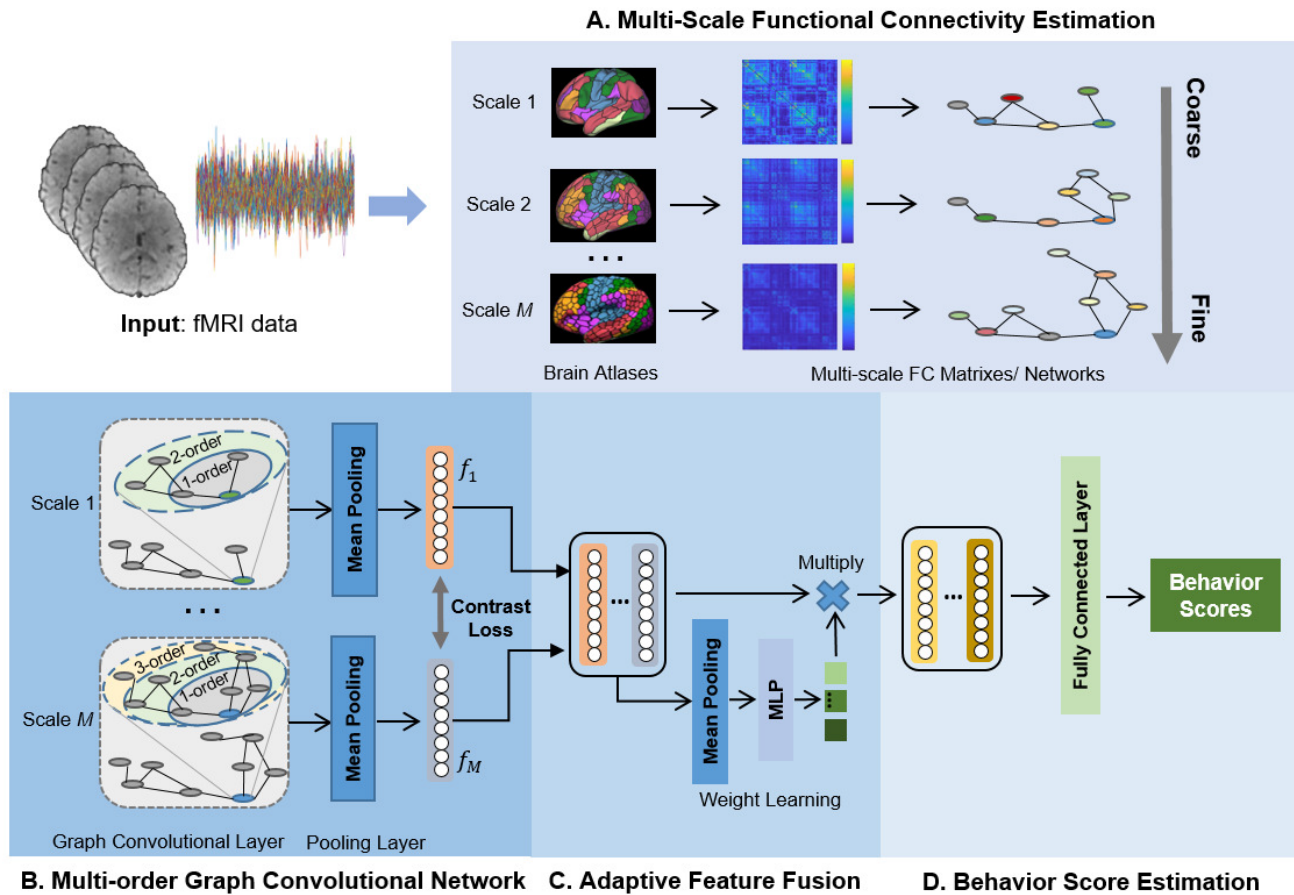


Fig. 1. The overall framework of MSFC-MO-GCN utilizes rs-fMRI data as input and outputs predicted behavioral scores, including four modules. (A) Based on rs-fMRI data, multiple FCNs are computed at different spatial scales by using multiscale brain parcellation atlas. (B) A multi-order graph convolutional network is used to learn brain connectivity representations at each scale. (C) An adaptive feature fusion module is used to integrate multi-scale FC features for each subject. (D) The joint features are fed into a fully connected layer for behavior prediction.

input and extracts the corresponding feature representation via a multi-order graph convolution layer and pooling layer. Our designed graph convolution layer considers the multi-order functional interactions of nodes, rather than only 1-order neighbors, to update nodal feature representations, enabling the use of high-order graph information for brain graph representation learning. Considering the potential redundancy in multi-scale FCNs, we add an inter-scale contrast constraint in GCN to improve the similarity of learned feature representations across different spatial scales for each individual. After obtaining multi-scale brain connectivity features, Step C fuses them using an attention block and obtains the joint feature representation. Finally, the behavior score of the subject is estimated based on a fully connected layer with the learned joint feature as input. The detailed description of each step is given below.

A. Multi-Scale Functional Connectivity Estimation

The main task of this module is to construct multi-scale FCNs based on rs-fMRI data, which serve as input for the subsequent modules to achieve comprehensive learning of brain connectivity representation. As shown in Fig. 1(A), we selected a set of coarse-to-fine brain parcellation atlases, and constructed corresponding FC matrices for each spatial

resolution, where each element in the matrix represents the Pearson correlation between the averaged BOLD signals of the two corresponding brain regions. Considering that the biological significance of negative connections is still unclear, we set all negative correlation values in each FC matrix to 0. Meanwhile, we performed sparsification on each FC matrix, retaining only the top 5% strongest edges to construct the brain functional network (or brain graph). In order to maintain the graph connectedness, for each node, we kept the top five strongest edges connected to them. Thus, given M coarse-to-fine brain parcellations, each subject will generate M FCNs representing the brain connectivity pattern at M spatial scales, denoted as $G = \{G^1, G^2, \dots, G^M\}$.

B. Multi-Order Graph Convolutional Network

With multi-scale FCNs as input, we utilized GCN to learn the feature representation of brain connectivity. For each subject n , we represent multi-scale FCNs as a graph set $G_n = \{G_n^1, G_n^2, \dots, G_n^M\}$. The graph at m -th scale is denoted as $G_n^m = (V_n^m, A_n^m, X_n^m)$, where V_n^m is the set of nodes at m -th scale, A_n^m is the adjacency matrix, and X_n^m is the attribute matrix initialized as identity matrix. The conventional GCNs extract deep feature representations of graph data by propagating node features through the graph Laplacian operator.

Despite being widely used, these methods have a limitation that ignores the rich information engendered by high-order functional interaction of brain regions because they only consider node's immediate neighbors. To tackle this problem, this paper proposed a novel GCN model for FC feature extraction, which consists of three important components: multi-order graph convolution layer, pooling layer, and inter-scale contrast constraint.

For the clear description in the following text, we first give the definition of multi-order neighbors. Specifically, given a node v , 0-order neighbor represents the vertex v itself; 1-order neighbors are the vertices adjacent to v (as shown in Fig. 1(B)); k -order neighbors is the set of vertices that can be reached from the node v by traversing exactly k edges (as shown in Fig. 1(B)).

1) **Multi-Order Graph Convolution Layer:** It can learn broader neighborhood relations by aggregating feature representations of neighbors at different distances as illustrated in Fig. 2. Specifically, given a graph G_n^m , let i -th layer's node feature be $(X_n^m)^i \in \mathbb{R}^{|V_n^m| \times d_i}$, where $|V_n^m|$ is the number of nodes and d_i is the feature dimension at i -th layer. In our proposed GCN model, the graph convolution operator is revised as

$$(X_n^m)^{(i+1)} = \parallel_{0 \leq k \leq K} \sigma((\tilde{A}_n^m)^k (X_n^m)^i W_k^i) \quad (1)$$

Here, $\tilde{A}_n^m = (\tilde{D}_n^m)^{-\frac{1}{2}} (A_n^m + I) (\tilde{D}_n^m)^{-\frac{1}{2}}$ is a symmetrically normalized adjacency matrix with self-connections, where \tilde{D}_n^m is the degree matrix of $(A_n^m + I)$. $k \in \{1, 2, \dots, K\}$ is the number of orders (i.e., steps) for the node to reach its neighbors. $(\tilde{A}_n^m)^k$ represents k -order adjacency matrix computed by multiplying \tilde{A}_n^m by k times; and when $k = 0$, $(\tilde{A}_n^m)^k$ is the identity matrix. $W_k^i \in \mathbb{R}^{d_i \times d_{i+1}}$ is the weight parameter matrix, σ is the ReLU activation function, \parallel means column-wise concatenation. Given a toy example with $K = 2$ in Fig. 2, the feature representation of the given node is updated by aggregating information from itself $((\tilde{A}_n^m)^0)$, 1-order neighboring nodes $((\tilde{A}_n^m)^1)$ and 2-order neighboring nodes $((\tilde{A}_n^m)^2)$ with column-wise concatenation. This method can extract effective features while keeping the graph structure unchanged. It is worth noting that, given the potential topological differences in the brain connectivity at different spatial scales, we allow for the use of different K values for FCNs with different resolutions.

2) **Pooling Layer:** After obtaining representation of each graph, we aggregated node features via the pooling layer and generated the final feature vector f_n^m for G_n^m . Specifically, given feature matrix X_n^m of subject n at scale m , f_n^m is computed as follows

$$f_n^m = p(X_n^m) = \frac{1}{|V_n^m|} \sum_{v=1}^{|V_n^m|} (x_v)_n^m \quad (2)$$

where $(x_v)_n^m$ represents the feature representation of node v of subject n at scale m .

3) **Inter-Scale Contrast Constraint:** Considering the potential information redundancy of FCs among different spatial scales, we designed an inter-scale contrastive loss L_{inter} to improve the within-subject across-scale similarity of graph features during the learning process. To this end, we treated feature

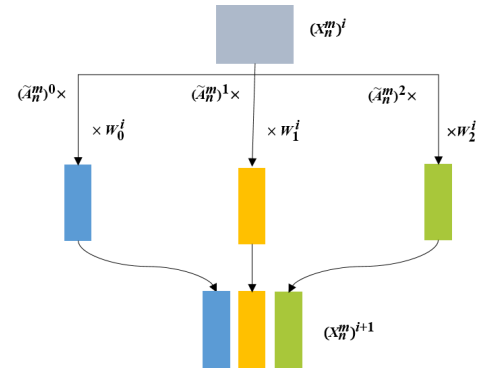


Fig. 2. Toy example of multi-order graph convolution layer with $K = 2$. In this figure, gray rectangle represents the input feature matrix, blue ones represent information from node itself, and yellow and green ones denotes connection from 1-order and 2-order neighbors, respectively.

representations of multi-scale FCNs from one subject as positive pairs, while those from different subjects as negative pairs, and formatted L_{inter} as follows:

$$L_{\text{inter}} = \sum_{m=1}^{M-1} \max \left(\text{Dist} \left(f_n^m, f_n^{m+1} \right) - \text{Dist} \left(f_n^m, f_s^m \right) + \delta, 0 \right) \quad (3)$$

where Dist function measures Euclidean distance between two vectors and δ is the margin parameter. (f_n^m, f_n^{m+1}) is the positive pair, representing the features learned from scale m and $m + 1$ of the same subject n ; (f_n^m, f_s^m) is the negative pair, representing the features learned from same spatial scale m of the subjects n and s and $n \neq s$. Equation (3) learns the correlation between multi-scale feature representations of subjects by minimizing the distance between positive pair representations and maximizing the distance between negative pair representations.

C. Adaptive Feature Fusion

After feature learning, we introduced an attention block to fuse features from multiple scales. This block was implemented by weight connection method with two steps. Given the subject n , let multi-scale feature representations be $f_n = \{f_n^1, f_n^2, \dots, f_n^M\}$ ($M = 3$). We first obtain the vector θ_n of scale-wise by global average pooling. Formally, the m -th element of θ_n is calculated by:

$$\theta_n^m = h_{\text{Avg}}(f_n^m) \quad (4)$$

where h_{Avg} means global average pooling. We then assessed the contribution weight φ_n of features with

$$\varphi_n = g(\theta_n) = \delta(Q\sigma(W\theta_n)) \quad (5)$$

where W and Q are trainable parameter, σ is ReLU function, and δ is sigmoid function. Finally, we concatenated multi-scale weighted features to generate joint features z_n of subject n as

$$z_n = \left\{ \varphi_n^1 f_n^1, \varphi_n^2 f_n^2, \dots, \varphi_n^M f_n^M \right\} \quad (6)$$

D. Behavior Score Estimation

Based on the learned joint features z_n , we estimated behavior score with fully connected layer as

$$\hat{Y} = z_n U, \quad (7)$$

where U represents a trainable parameter. Finally, two supervised loss terms were used for the behavior prediction over all training subjects defined as

$$L_{\text{total}} = \alpha L_{\text{inter}} + L_{\text{MAE}} \quad (8)$$

The first term is inter-scale contrastive loss defined in Equation (3). The second term measures the accuracy of the prediction model formatted as

$$L_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N E(y_i, \hat{y}_i), \quad (9)$$

where E represents the absolute error between the real and predicted behavior scores and N is the number of training samples. α is a hyperparameter used to balance contributions of two loss terms.

V. IMPLEMENTATION

A. Model Settings and Evaluation Metric

In this paper, we implemented MSFC-MO-GCN based on Tensorflow using the Python language. The model consists of two multi-order graph convolutional layers, with 96 and 12 filters respectively, as well as a 12-channel pooling layer and a 1-channel fully connected layer. We used Adam as the optimizer with a learning rate of 0.005, and applied L2 regularization of 0.0005 to control overfitting and noise in imaging data. The network was trained on 805 subjects using Adam optimizer, with a 5-fold cross-validation strategy, a batch size of 16, and 70 iterations. Specifically, we randomly divided the 805 subjects into 5 folds and used internal cross-validation to determine the model's hyperparameters. We then evaluated the model's performance on the test set by calculating the Pearson correlation between predicted and actual scores of behaviors. To ensure the stability of the results, we repeated the 5-fold cross-validation 5 times for each experiment, and used the average result from the 25 folds as the final prediction accuracy. All experiments were accelerated on an Nvidia RTX 2080 GPU.

B. Compared Methods

Eight popular FC-based behavior prediction models were used for comparison in this paper, including kernel regression method [54], FNN [55], BrainNetCNN [56], GCNN [11], GAT [57], SAGPool [58], Meta-RegGNN [59], and BC-GCN-SE [60].

1) *Kernel Regression Method*: it is the conventional machine learning based individual behavior prediction model. This method first calculates the similarity between subjects based on FCs, and then predicts the behavior score of the test subject by taking the weighted average of the behavioral measures of all training subjects. Ridge regression is adopted in this model.

2) *FNN*: it belongs to a class of feedforward neural networks. This method treats FCN as a vector and adopts several fully connected layers to achieve behavior score prediction.

3) *BrainNetCNN*: it is a specially designed deep neural network for brain connectivity network. It takes in FCN directly as input and outputs behavior prediction score, which consists of four types of layers: Edge-to-Edge (E2E) layer, Edge-to-Node (E2N) layer, Node-to-Graph (N2G) layer and a final fully connected layer.

4) *GCNN*: it is the most conventional GCN method. It takes in vectorized FCNs of all subjects as input and outputs behavior scores of all subjects. The graph in this model is not brain network but individual similarity matrix.

5) *GAT*: it is a classic attention-based GCN model that leverages an attention mechanism to assess the influence of adjacent nodes in a graph, facilitating dynamic feature learning tailored to each node's network context. It processes FCN of subjects to predict behavioral scores.

6) *SAGPool*: it employs a self-attention mechanism for dynamic downsampling of graphs, selectively focusing on nodes most important to the task, and thus enhances efficiency and effectiveness in GNN processing. It takes in FCN of each subject as input and generates behavior prediction scores.

7) *Meta-RegGNN*: it is a meta-learning regression GNN model specifically designed for FC-based behavior prediction. It inputs FCN of subjects and outputs behavior prediction scores.

8) *BC-GCN-SE*: it is an edge-based graph path convolution method that adeptly aggregates information across different paths, making it suitable for densely connected brain graphs. The model includes Graph Path Convolution, Edge Pooling, and Node Pooling layers and processes FCN of subjects to produce behavior prediction scores.

We implemented kernel regression method using the Matlab platform, and constructed seven deep learning methods based on Tensorflow using Python language. The model settings and training methods were the same as those for MSFC-MO-GCN. The difference is that the comparative methods uses L_{MAE} as the loss function, adopts the single-scale FCN as input, and has different settings of layers and units, as summarized in Table I. All methods chose FCN with 500 ROIs as the input for model training as this network scale achieves the highest behavioral prediction accuracy.

VI. RESULTS AND DISCUSSION

In this section, we first present the analysis results of the relationship between FCs and behaviors at different spatial scales. We then conduct a comprehensive evaluation of the proposed individual behavior prediction model, including the parameter analysis, comparison with four conventional methods, and ablation experiments. Finally, we discuss the importance of functional subsystems in behavior prediction to achieve interpretability of the features in the model.

A. Comparison of FC-Behavior Relationship Between Different Scales

Fig. 3 presents the results of brain-behavior relationships at the system level across three spatial scales. The comparison

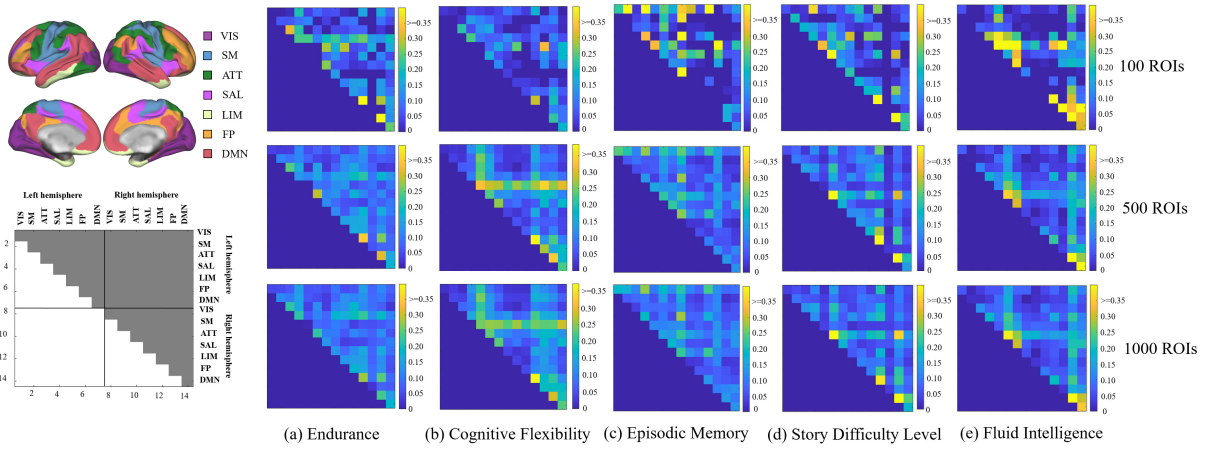


Fig. 3. Different CSs of intra- and inter-system connections with five behavior metrics at three spatial scales, including (a) endurance, (b) cognitive flexibility, (c) episodic memory, (d) story difficulty level and (e) fluid intelligence. Networks 1-7 and 8-14 represent the seven functional subsystems in the left and right hemispheres, respectively. Only the upper triangle of the connectivity matrix is presented since the matrix is symmetric.

TABLE I
PARAMETER SETTINGS OF COMPARATIVE METHODS

Methods	FCN	Number of Layers and Units
Kernel Regression	FCN-500	–
FNN	FCN-500	4{32, 18, 8, 1}
BrainNetCNN	FCN-500	4{16, 128, 26, 1}
GCNN	FCN-500	2{32, 1}
GAT	FCN-500	3{256, 1, 1}
SAGPool	FCN-500	6{32, 32, 32, 32, 16, 1}
Meta-RegGNN	FCN-500	3{64, 1, 1}
BC-GCN-SE	FCN-500	6{16, 16, 16, 64, 256, 1}

of each column in Fig. 3 reveals significant differences in the correlation between FC and behavior across various scales. This finding validates our previous hypothesis: the human brain has a hierarchical structure, with brain-behavior relationships varying across different spatial scales, highlighting the significance of employing multi-scale functional connectivity (FC) in behavior prediction. We also noted that the differences between 100 ROIs and 500 ROIs are generally more significant than those between 500 ROIs and 1000 ROIs, especially in tasks related to cognitive flexibility, episodic memory, and fluid intelligence. This result implies that as spatial resolution increases, the connectivity information from different scales of functional networks might overlap, leading to redundancy. This highlights the necessity for incorporating inter-scale contrast learning in multi-scale FC-based behavior prediction models.

Moreover, by comparing each row in Fig. 3, we observed significant variations in the patterns of FC-behavior relationships across various behavioral tasks. For example, as shown in Fig. 3b, in the cognitive flexibility task at ROI = 500, both intra- and inter-system CS values are high in LIM system, while the similar result is not observed in the other four cognitive tasks. This implies a potential involvement of the LIM system in the brain's cognitive flexibility processing. These observations highlight the limitations of traditional machine

learning methods, which often depend on expert-driven feature extraction in behavior prediction, and underscore the significance of leveraging deep learning techniques for automated feature extraction.

B. Parameter Analysis

There are two hyper-parameters in MSFC-MO-GCN that may affect the accuracy of individual behavior prediction. One is α in the loss function, which controls the contribution of inter-scale contrastive constraints, and the other is the maximum order K in the multi-order graph convolution layer. In this subsection, we used the method of controlling variables to evaluate the impact of these two parameters on the prediction results. Similar to the previous experiments, the average Pearson correlation coefficient in 25 test folds was used as the evaluation metric. Since different scale FCNs can use different K values, in this paper, $K = (K_{FCN-100}, K_{FCN-500}, K_{FCN-1000})$, representing the maximum order of graph convolution under three spatial scales of ROI = 100, ROI = 500, and ROI = 1000, respectively.

1) *Parameter α* : In this experiment, we fixed the K to five different combinations, i.e., {2, 2, 1}, {3, 1, 1}, {2, 1, 1}, {1, 1, 1}, and {3, 1, 1}. For each combination, we varied α within the range of {0, 0.01, 0.1, 1, 10} and calculated the prediction accuracy of MSFC-MO-GCN for each setting, as shown in Fig. 4. We can observe that the performance of MSFC-MO-GCN is significantly affected by the parameter α , exhibiting a general trend of increasing first and then decreasing. Specifically, the correlation value increases rapidly from 0 to 0.01, fluctuates slightly between 0.01 and 1 (first decreasing then increasing generally), and then sharply decreases from 1 to 10. These results of MSFC-MO-GCN performance varies as α demonstrate the necessity of introducing inter-scale contrastive constraints. Furthermore, the relatively smooth curve from 0.01 to 1 compared to other ranges indicates that the L_{MAE} loss plays a more important role than L_{Inter} in the feature learning process. The significant drop in performance from 1 to 10 suggests that highly similar FCN features across various scales of the same subject may lead to feature redundancy, which subsequently results in

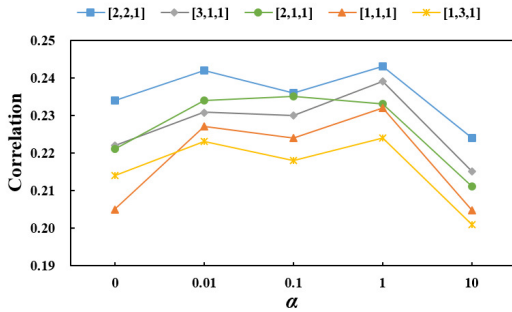


Fig. 4. Parameter analysis results of parameter α . The line in different color represents different settings of K , where the blue one represents $\{2, 2, 1\}$, gray line represents $\{3, 1, 1\}$, green line represents $\{2, 1, 1\}$, orange line represents $\{1, 1, 1\}$ and yellow one represents $\{1, 3, 1\}$.

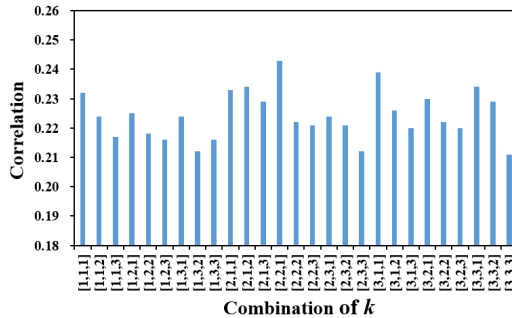


Fig. 5. Parameter analysis of the number of orders K . The settings of K were varied from 1 to 3 at each spatial scale, thus generating 27 combinations.

decreased model performance. In the subsequent experiments, α was set to 1.

2) *Parameter K*: In this experiment, we fixed α to 1, and varied the maximum accessible order of GCN K at each scale from 1 to 3. As a result, 27 combinations ($3 \times 3 \times 3$) of FCNs were generated across three spatial scales (i.e., ROI = 100, 500 and 1000). For each combination, we evaluated the individual behavioral prediction performance of MSFC-MO-GCN, and presented the results as bar charts in Fig. 5. We observe that the individual behavioral prediction ability of MSFC-MO-GCN is significantly influenced by the parameter K , demonstrating the necessity of using multi-order connection information in the graph convolution process. Through further analysis, we find that the top three performing GCN models have K values of $\{2, 2, 1\}$, $\{3, 1, 1\}$, and $\{3, 3, 1\}$, while the three worst performing models have K values of $\{3, 3, 3\}$, $\{1, 3, 2\}$, and $\{2, 3, 3\}$. We compare these two sets of parameter settings and are amazed to discover that the smaller FC networks use larger K values, whereas the larger-scale FC networks use smaller K values, having more potential to achieve the superior model accuracy in predicting behavior. This may be because each node in the smaller scale FCN contains larger brain regions, greatly increasing the probability of the transmission of multi-order connection information between nodes. In the following experiments, K was set to $\{2, 2, 1\}$.

C. Comparison With Other Methods

We compared the performance of our proposed method with kernel regression method, FNN, BrainNetCNN, GCNN,

GAT, SAGPool, Meta-RegGNN, and BC-GCN-SE. The implementation details and parameter setting of all methods were illustrated in Section V. In MSFC-MO-GCN, $\alpha = 1$ and $K = \{2, 2, 1\}$. Similarly, the average of 25 folds was thus used as the final prediction accuracy. Table II. provides the comparison results between different models in predicting five types of human behaviors.

As shown in Table II, we can see that our proposed method significantly outperforms other comparative methods. This not only confirms the superiority of our approach but also indicates that extracting multi-scale and multi-order FC information can yield a more comprehensive representation of brain connectivity, thereby enhancing the accuracy of behavioral predictions. Among the eight comparative methods, SAGPool performed best in the endurance and cognitive flexibility prediction tasks, BC-GCN-SE had the highest accuracy in predicting episodic memory and fluid intelligence, and meta-RegGNN excelled in the story difficulty level task. In contrast, GCNN performed worst in all behavioral prediction tasks, possibly because it extracts features and calculates behavioral scores based on individual similarities rather than FC information. Moreover, we found that traditional kernel regression methods performed comparably to FNN, BrainNetCNN, and GAT, for example, in predicting endurance, cognitive flexibility, episodic memory, and fluid intelligence. This may be because the sample size in this study was insufficient to fully meet the training requirements of deep learning methods. Furthermore, this result implies that existing deep learning methods cannot effectively and comprehensively extract FC feature information.

From the differences in the model's results on different behavior prediction tasks, we found that the behavioral prediction stability of MSFC-MO-GCN is significantly higher than that of other compared methods. For example, FNN has much lower prediction accuracy on cognitive flexibility and episodic memory than the other two behaviors; the kernel regression model has significantly lower prediction accuracy on episodic memory than the other behaviors. This result suggests that, our proposed model has better task generalization ability than the other methods, which may be due to the richer FC information used in the MSFC-MO-GCN model, making it easier to extract features related to various human behaviors.

D. Ablation Study

To demonstrate the effectiveness of introducing multi-scale FCs, multi-order graph convolution, and inter-scale contrast constraint, we designed three baseline models for comparison with MSFC-MO-GCN in fluid intelligence prediction task. The average of Pearson's correlation coefficient on 25 folds was used as the behavioral prediction evaluation index.

1) *Baseline 1 (Single-Scale FCs + Multi-Order Graph Convolution)*: The model takes in single-scale FCNs as input and uses multi-order graph convolution to learn the corresponding feature representation. At each scale of FCN, we evaluated the model performance with K equals to 1, 2 and 3, respectively.

2) *Baseline 2 (Multi-Scale FCs + 1-Order Graph Convolution + Inter-Scale Contrast Constraint)*: It takes in multi-scale FCNs as input and uses traditional graph convolution method (i.e., 1-order graph convolution) for the feature learning.

TABLE II
COMPARISON OF BEHAVIORAL PREDICTION PERFORMANCE

Method	Endurance	Cognitive Flexibility	Episodic Memory	Story Difficulty Level	Fluid Intelligence
Kernel	0.160±0.062	0.141±0.068	0.085±0.068	0.151±0.069	0.163±0.048
BrainnetCNN	0.132±0.057	0.112±0.039	0.090±0.066	0.133±0.065	0.123±0.061
FNN	0.163±0.070	0.070±0.060	0.060±0.051	0.167±0.070	0.170±0.058
GCNN	0.018±0.076	0.020±0.050	0.034±0.065	0.036±0.060	0.050±0.079
GAT	0.064±0.063	0.094±0.054	0.094±0.053	0.160±0.048	0.145±0.071
SAGPool	0.165±0.081	0.167±0.034	0.101±0.088	0.177±0.063	0.159±0.052
meta-RegGNN	0.124±0.037	0.104±0.046	0.092±0.062	0.186±0.044	0.169±0.048
BC-GCN-SE	0.096±0.069	0.135±0.056	0.112±0.063	0.176±0.049	0.192±0.037
MSFC-MO-GCN	0.193±0.028	0.174±0.026	0.184±0.052	0.205±0.034	0.243±0.038

Inter-scale contrast constraint was also added in this model. In this experiment, α was set to 1, and K was set to $\{1, 1, 1\}$.

3) *Baseline 3 (Multi-Scale FCs + Multi-Order Graph Convolution)*: It takes in multi-scale FCNs as input and uses multi-order graph convolution as convolutional layer, but removes inter-scale contrast constraint during feature learning. We set α to 1, and reported the best prediction result with $K = \{2, 2, 1\}$.

The comparison results were presented in Table III. It can be seen that the prediction accuracy of the three baseline models is lower than that of MSFC-MO-GCN, indicating that each proposed strategy contributes to improving the model's predictive ability. Further comparison reveals that the abilities of baseline model 2 and 3 are both superior to baseline model 1, suggesting that the introduction of multi-scale FC has the greatest contribution to improving the predictive performance of individual behavior. More interestingly, by comparing the prediction accuracies based on FCs at three different scales in baseline model 1, we found that the medium-sized FCN (ROI = 500) is more likely to achieve better performance. This may be because small-scale FCNs cannot obtain detailed functional interactions between brain regions, while large-scale FCNs are prone to false connections introduced by imaging noise.

E. Importance of Functional Connectivity

In this subsection, we discussed the importance of FC for MSFC-MO-GCN in each behavior prediction task. In this paper, we used occlusion importance (OI) [61] as the evaluation metric. Specifically, given the FC to be evaluated, we estimated its OI by computing the absolute difference of the predicted behavior scores based on FCNs before and after removing this edge. To reduce computational complexity, the FC was classified into seven functional subsystems, i.e., VIS, SM, ATT, SAL, LIM, FP, and DMN, and the importance of FCs within each subsystem was calculated for each behavior prediction task. We computed the OI value based on test results on 25 folds.

Fig. 6 provides a detailed display of the OI results for various functional subsystems across five different behavioral prediction tasks. In this analysis, we observed that

TABLE III
COMPARISON RESULTS OF MSFC-MO-GCN WITH THREE BASELINE MODELS

Models	Input Data	K	Prediction Accuracy (r)
Baseline1	FCN-100	1	0.196 ± 0.053
		2	0.204 ± 0.050
		3	0.201 ± 0.055
	FCN-500	1	0.202 ± 0.031
		2	0.206 ± 0.039
		3	0.214 ± 0.045
	FCN-1000	1	0.181 ± 0.031
		2	0.191 ± 0.034
		3	0.192 ± 0.037
Baseline2	FCN-100, FCN-500, FCN-1000	$\{1, 1, 1\}$	0.232 ± 0.039
Baseline3	FCN-100, FCN-500, FCN-1000	$\{2, 2, 1\}$	0.234 ± 0.043
MSFC-MO-GCN	FCN-100, FCN-500, FCN-1000	$\{2, 2, 1\}$	0.243 ± 0.038

the influential functional connections impacting behavior are distributed across distinct functional subsystems in different tasks. For example, in endurance prediction tasks related to physical activity, the sensorimotor (SM) system, closely tied to movement control and bodily coordination, exhibited the highest OI values, aligning with expectations [62]. In cognitive flexibility tasks, systems requiring heightened attention and information processing, such as the ATT [63], FP [64], and DMN [65], played equally crucial roles. Conversely, in fluid intelligence task, the differences among the seven functional subsystems were minimal, reflecting the broad brain area involvement these tasks demand [66]. These findings suggest that the execution of higher-level cognitive functions tend to depend on more complex and extensive neural networks. From endurance to fluid intelligence, there is an apparent increase in the number of functional subsystems involved in task processing.

Additionally, our analysis also found that the DMN consistently demonstrated relatively high OI values across all five behavioral tasks. This could be attributed to the fact that brain regions within the DMN, such as the prefrontal cortex,

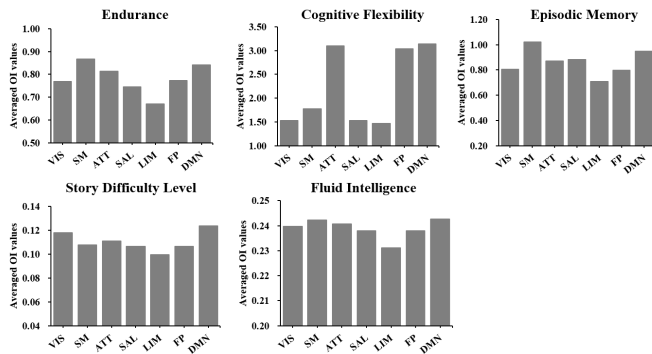


Fig. 6. Importance evaluation results of FCs for each behavior prediction.

anterior cingulate cortex, and posterior cingulate cortex, serve as key hubs for information transmission and integration during various behavioral and cognitive activities. This discovery underscores the widespread importance of the DMN across different types of behavioral and cognitive tasks and its critical role in understanding how the brain coordinates and executes complex tasks.

VII. CONCLUSION

This paper proposes a multi-order graph convolutional network model for individual behavior prediction based on multi-scale functional connectivity and GCN, named as MSFC-MO-GCN. In this model, we consider the hierarchical structure of brain system and uses FCs from multiple spatial scales for brain connectivity representation learning. We also designs multi-order graph convolution layer to extract information from far-distance functional interactions to enrich feature set learned from nodal connections. By introducing these two strategies, our newly proposed GCN model has larger ability to learn the brain-behavior relationship representation. Experimental results on a publicly available dataset from human connection project show that MSFC-MO-GCN exhibits better performance compared to competing methods.

ACKNOWLEDGMENT

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The patients/participants provided their written informed consent to participate in this study. The authors have no conflicts of interest to declare.

REFERENCES

- [1] X. Fan and H. Markram, "A brief history of simulation neuroscience," *Frontiers Neuroinform.*, vol. 13, p. 32, May 2019.
- [2] C. Eliasmith and O. Trujillo, "The use and abuse of large-scale brain models," *Current Opinion Neurobiol.*, vol. 25, pp. 1–6, Apr. 2014.
- [3] M. D. Rosenberg, B. J. Casey, and A. J. Holmes, "Prediction complements explanation in understanding the developing brain," *Nature Commun.*, vol. 9, no. 1, p. 589, Feb. 2018.
- [4] N. Kriegeskorte and P. K. Douglas, "Cognitive computational neuroscience," *Nature Neurosci.*, vol. 21, no. 9, pp. 1148–1160, 2018.
- [5] R. Kanai and G. Rees, "The structural basis of inter-individual differences in human behaviour and cognition," *Nature Rev. Neurosci.*, vol. 12, no. 4, pp. 231–242, Apr. 2011.

- [6] A. Llera, T. Wolfers, P. Mulders, and C. F. Beckmann, "Inter-individual differences in human brain structure and morphology link to variation in demographics and behavior," *eLife*, vol. 8, Jul. 2019, Art. no. e44443.
- [7] P. Shah et al., "Mapping the structural and functional network architecture of the medial temporal lobe using 7T MRI," *Hum. Brain Mapping*, vol. 39, no. 2, pp. 851–865, Feb. 2018.
- [8] S. Mueller et al., "Individual variability in functional connectivity architecture of the human brain," *Neuron*, vol. 77, no. 3, pp. 586–595, Feb. 2013.
- [9] N. P. Friedman and A. Miyake, "Unity and diversity of executive functions: Individual differences as a window on cognitive structure," *Cortex*, vol. 86, pp. 186–204, Jan. 2017.
- [10] W. W. Seeley et al., "Dissociable intrinsic connectivity networks for salience processing and executive control," *J. Neurosci.*, vol. 27, no. 9, pp. 2349–2356, Feb. 2007.
- [11] T. He et al., "Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics," *NeuroImage*, vol. 206, Feb. 2020, Art. no. 116276.
- [12] X. Shen et al., "Using connectome-based predictive modeling to predict individual behavior from brain connectivity," *Nature Protocols*, vol. 12, no. 3, pp. 506–518, Mar. 2017.
- [13] M. Li et al., "Performing group-level functional image analyses based on homologous functional regions mapped in individuals," *PLOS Biol.*, vol. 17, no. 3, Mar. 2019, Art. no. e2007032.
- [14] R. Jiang et al., "Gender differences in connectome-based predictions of individualized intelligence quotient and sub-domain scores," *Cerebral Cortex*, vol. 30, no. 3, pp. 888–900, Mar. 2020.
- [15] L. Steardo et al., "Application of support vector machine on fMRI data as biomarkers in schizophrenia diagnosis: A systematic review," *Frontiers Psychiatry*, vol. 11, p. 588, Jun. 2020.
- [16] J. Sui, R. Jiang, J. Bustillo, and V. Calhoun, "Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: Methods and promises," *Biol. Psychiatry*, vol. 88, no. 11, pp. 818–828, Dec. 2020.
- [17] G. Orrù, W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review," *Neurosci. Biobehav. Rev.*, vol. 36, no. 4, pp. 1140–1152, Apr. 2012.
- [18] Y. Liang, B. Liu, and H. Zhang, "A convolutional neural network combined with prototype learning framework for brain functional network classification of autism spectrum disorder," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 2193–2202, 2021.
- [19] O. J. Arthurs and S. Boniface, "How well do we understand the neural origins of the fMRI BOLD signal?" *Trends Neurosci.*, vol. 25, no. 1, pp. 27–31, Jan. 2002.
- [20] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nature Rev. Neurosci.*, vol. 8, no. 9, pp. 700–711, Sep. 2007.
- [21] A. B. Leber, N. B. Turk-Browne, and M. M. Chun, "Neural predictors of moment-to-moment fluctuations in cognitive flexibility," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 36, pp. 13592–13597, Sep. 2008.
- [22] A. Zhang et al., "Joint Bayesian-incorporating estimation of multiple Gaussian graphical models to study brain connectivity development in adolescence," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 357–365, Feb. 2020.
- [23] R. Li, G. Rui, C. Zhao, C. Wang, F. Fang, and Y. Zhang, "Functional network alterations in patients with amnesic mild cognitive impairment characterized using functional near-infrared spectroscopy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 123–132, Jan. 2020.
- [24] E. S. Finn et al., "Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity," *Nature Neurosci.*, vol. 18, no. 11, pp. 1664–1671, Nov. 2015.
- [25] R. N. Spreng, A. W. Lockrow, E. DuPre, R. Setton, K. A. Spreng, and G. R. Turner, "Semanticized autobiographical memory and the default-executive coupling hypothesis of aging," *Neuropsychologia*, vol. 110, pp. 37–43, Feb. 2018.
- [26] R. Kong et al., "Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion," *Cerebral Cortex*, vol. 29, no. 6, pp. 2533–2551, Jun. 2019.
- [27] J.-Y. Yun and Y.-K. Kim, "Graph theory approach for the structural-functional brain connectome of depression," *Prog. Neuro-Psychopharmacol. Biol. Psychiatry*, vol. 111, Dec. 2021, Art. no. 110401.
- [28] G. Pan, L. Xiao, Y. Bai, T. W. Wilson, J. M. Stephen, V. D. Calhoun, and Y.-P. Wang, "Multiview diffusion map improves prediction of fluid intelligence with two paradigms of fMRI analysis," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 8, pp. 2529–2539, Aug. 2021.

- [29] M. W. Cole, T. Yarkoni, G. Repovs, A. Anticevic, and T. S. Braver, "Global connectivity of prefrontal cortex predicts cognitive control and intelligence," *J. Neurosci.*, vol. 32, no. 26, pp. 8988–8999, Jun. 2012.
- [30] J. S. Siegel et al., "Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 30, pp. E4367–E4376, Jul. 2016.
- [31] M. D. Rosenberg, E. S. Finn, D. Scheinost, R. T. Constable, and M. M. Chun, "Characterizing attention with predictive network models," *Trends Cognit. Sci.*, vol. 21, no. 4, pp. 290–302, Apr. 2017.
- [32] M. D. Rosenberg et al., "A neuromarker of sustained attention from whole-brain functional connectivity," *Nature Neurosci.*, vol. 19, no. 1, pp. 165–171, Jan. 2016.
- [33] M. E. Magnuson et al., "Errors on interrupter tasks presented during spatial and verbal working memory performance are linearly linked to large-scale functional network connectivity in high temporal resolution resting state fMRI," *Brain Imag. Behav.*, vol. 9, no. 4, pp. 854–867, Dec. 2015.
- [34] A. D. Nostro et al., "Predicting personality from network-based resting-state functional connectivity," *Brain Struct. Function*, vol. 223, no. 6, pp. 2699–2719, Jul. 2018.
- [35] Z. Cui and G. Gong, "The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features," *NeuroImage*, vol. 178, pp. 622–637, Sep. 2018.
- [36] S. Mansour L, Y. Tian, B. T. T. Yeo, V. Cropley, and A. Zalesky, "High-resolution connectomic fingerprints: Mapping neural identity and behavior," *NeuroImage*, vol. 229, Apr. 2021, Art. no. 117695.
- [37] U. Pervaiz, D. Vidaurre, M. W. Woolrich, and S. M. Smith, "Optimising network modelling methods for fMRI," *NeuroImage*, vol. 211, May 2020, Art. no. 116604.
- [38] T. He et al., "Is deep learning better than kernel regression for functional connectivity prediction of fluid intelligence?" in *Proc. Int. Workshop Pattern Recognit. Neuroimag. (PRNI)*, Singapore, Jun. 2018, pp. 1–4.
- [39] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–14.
- [40] M. Li, Z. Ma, Y. G. Wang, and X. Zhuang, "Fast Haar transforms for graph neural networks," *Neural Netw.*, vol. 128, pp. 188–198, Aug. 2020.
- [41] Y. G. Wang, M. Li, Z. Ma, G. Montufar, X. Zhuang, and Y. Fan, "Haar graph pooling," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 9952–9962.
- [42] X. Zheng, B. Zhou, M. Li, Y. G. Wang, and J. Gao, "MathNet: Haar-like wavelet multiresolution analysis for graph representation learning," *Knowl.-Based Syst.*, vol. 273, Aug. 2023, Art. no. 110609.
- [43] A. Schaefer et al., "Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI," *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, Sep. 2018.
- [44] R. Wang, P. Lin, M. Liu, Y. Wu, T. Zhou, and C. Zhou, "Hierarchical connectome modes and critical state jointly maximize human brain functional diversity," *Phys. Rev. Lett.*, vol. 123, no. 3, Jul. 2019, Art. no. 038301.
- [45] M. Liu et al., "Multiscale functional connectome abnormality predicts cognitive outcomes in subcortical ischemic vascular disease," *Cerebral Cortex*, vol. 32, no. 21, pp. 4641–4656, Oct. 2022.
- [46] S. Abu-El-Haija et al., "MixHop: Higher-order graph convolutional architectures via sparsified neighborhood mixing," in *Proc. Int. Conf. Mach. Learn.*, San Francisco, CA, USA, 2019, pp. 21–29.
- [47] T. Xie, B. Wang, and C.-C. J. Kuo, "GraphHop: An enhanced label propagation method for node classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 92870–9301, Nov. 2023.
- [48] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil, "The WU-minn human connectome project: An overview," *NeuroImage*, vol. 80, pp. 62–79, Oct. 2013.
- [49] K. B. Casaletto et al., "Demographically corrected normative standards for the English version of the NIH toolbox cognition battery," *J. Int. Neuropsychol. Soc.*, vol. 21, no. 5, pp. 378–391, May 2015.
- [50] M. F. Glasser et al., "The minimal preprocessing pipelines for the human connectome project," *NeuroImage*, vol. 80, pp. 105–124, Oct. 2013.
- [51] L. Griffanti et al., "ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging," *NeuroImage*, vol. 95, pp. 232–247, Jul. 2014.
- [52] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith, "Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers," *NeuroImage*, vol. 90, pp. 449–468, Apr. 2014.
- [53] B. T. Thomas Yeo et al., "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *J. Neurophysiol.*, vol. 106, no. 3, pp. 1125–1165, Sep. 2011.
- [54] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
- [55] E. Akleman, "Deep learning," *Computer*, vol. 53, no. 9, p. 17, Sep. 2020.
- [56] J. Kawahara et al., "BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017.
- [57] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.
- [58] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3734–3743.
- [59] I. Jegham and I. Rekek, "Meta-RegGNN: Predicting verbal and full-scale intelligence scores using graph neural networks and meta-learning," in *Proc. Int. Workshop Predictive Intell. Med.* Cham, Switzerland: Springer, 2022, pp. 203–211.
- [60] Y. Li et al., "Brain connectivity based graph convolutional networks and its application to infant age prediction," *IEEE Trans. Med. Imag.*, vol. 41, no. 10, pp. 2764–2776, Oct. 2022.
- [61] L. Faber, K. A. Moghaddam, and R. Wattenhofer, "Contrastive graph neural network explanation," in *Proc. 37th Graph Represent. Learn. Beyond Workshop Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 1–6.
- [62] J. Borghuis, A. L. Hof, and K. A. Lemmink, "The importance of sensory-motor control in providing core stability: Implications for measurement and training," *Sports Med.*, vol. 38, pp. 893–916, Nov. 2008.
- [63] Z. Zhang et al., "Functional degeneration in dorsal and ventral attention systems in amnesic mild cognitive impairment and Alzheimer's disease: An fMRI study," *Neurosci. Lett.*, vol. 585, pp. 160–165, Jan. 2015.
- [64] J. L. Vincent, I. Kahn, A. Z. Snyder, M. E. Raichle, and R. L. Buckner, "Evidence for a frontoparietal control system revealed by intrinsic functional connectivity," *J. Neurophysiol.*, vol. 100, no. 6, pp. 3328–3342, Dec. 2008.
- [65] J. Smallwood, B. C. Bernhardt, R. Leech, D. Bzdok, E. Jefferies, and D. S. Margulies, "The default mode network in cognition: A topographical perspective," *Nature Rev. Neurosci.*, vol. 22, no. 8, pp. 503–513, Aug. 2021.
- [66] C. M. Clark, L. Lawlor-Savage, and V. M. Goghari, "Comparing brain activations associated with working memory and fluid intelligence," *Intelligence*, vol. 63, pp. 66–77, Jul. 2017.