Neural Networks 164 (2023) 91–104

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

# A deep connectome learning network using graph convolution for connectome-disease association study

Yanwu Yang [a,b], Chenfei Ye [b,c,1], Ting Ma [a,b,c,d,*]

[a] Department of Electronic and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, China
[b] Peng Cheng Laboratory, Shenzhen, China
[c] International Research Institute for Artificial Intelligence, Harbin Institute of Technology at Shenzhen, Shenzhen, China
[d] Guangdong Provincial Key Laboratory of Aerospace Communication and Networking Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China

ARTICLE INFO

*Article history:*
Received 20 September 2022
Received in revised form 1 February 2023
Accepted 16 April 2023
Available online 22 April 2023

Dataset link: https://github.com/podismine/MDCN

*Keywords:*
Deep connectome learning
Distance-based connectome network
Connectome-wide association study
Graph neural network

ABSTRACT

Multivariate analysis approaches provide insights into the identification of phenotype associations in brain connectome data. In recent years, deep learning methods including convolutional neural network (CNN) and graph neural network (GNN), have shifted the development of connectome-wide association studies (CWAS) and made breakthroughs for connectome representation learning by leveraging deep embedded features. However, most existing studies remain limited by potentially ignoring the exploration of region-specific features, which play a key role in distinguishing brain disorders with high intra-class variations, such as autism spectrum disorder (ASD), and attention deficit hyperactivity disorder (ADHD). Here, we propose a multivariate distance-based connectome network (MDCN) that addresses the local specificity problem by efficient parcellation-wise learning, as well as associating population and parcellation dependencies to map individual differences. The approach incorporating an explainable method, parcellation-wise gradient and class activation map (p-GradCAM), is feasible for identifying individual patterns of interest and pinpointing connectome associations with diseases. We demonstrate the utility of our method on two largely aggregated multicenter public datasets by distinguishing ASD and ADHD from healthy controls and assessing their associations with underlying diseases. Extensive experiments have demonstrated the superiority of MDCN in classification and interpretation, where MDCN outperformed competitive state-of-the-art methods and achieved a high proportion of overlap with previous findings. As a CWAS-guided deep learning method, our proposed MDCN framework may narrow the bridge between deep learning and CWAS approaches, and provide new insights for connectome-wide association studies.



## 1. Introduction

The drive towards decoding functional activity and neural communication of the brain network, or the connectome (Sporns, Tononi, & Kötter, 2005), enables us to understand the complex relationship between individual brain network organization and behavioral phenotypes (Bargmann & Marder, 2013; Bullmore & Bassett, 2011). Promising progress has been made in the past decade using neuroimaging techniques to identify brain network alterations underlying brain disorders (Fornito, Zalesky, & Breakspear, 2015; van den Heuvel, Scholtens, & Kahn, 2019; van den Heuvel & Sporns, 2019). To capture a rich set of connectome-phenotype associations, the concept of connectome-wide association studies (CWAS) suggests examining each brain connectivity variation from the perspective of the global network level, rather than in isolation (Shehzad et al., 2014). Recent CWAS investigations have successfully identified functional disconnectivity in neurodegenerative and psychiatric brain diseases (Sharma et al., 2017; Yang et al., 2021; Ye, Mori, Chan, & Ma, 2019).

In practice, two main analytical strategies have been widely used under the CWAS framework. Network-based statistical (NBS) analysis (Zalesky, Fornito, & Bullmore, 2010) was the first proposed CWAS method to avoid the multiple comparisons problems caused by the massive univariate generalized linear model (GLM) on each inter-voxel or inter-regional connectivity (Bellec et al., 2015). Another strategy is multivariate distance matrix regression (MDMR) (Milham, 2012; Shehzad et al., 2014). By estimating the

* Correspondence to: Department of Electronics and Information, Harbin Institute of Technology at Shenzhen, Rm 1206, Information Building, HIT Campus, Shenzhen University Town, Nanshan District, Shenzhen, Guangdong Province, 518055, China.
*E-mail addresses:* 20b952019@stu.hit.edu.cn (Y. Yang), yechenfei@hit.edu.cn (C. Ye), tma@hit.edu.cn (T. Ma).
[1] Co-first author.

https://doi.org/10.1016/j.neunet.2023.04.025

similarity distance of brain connectivity vectors among individuals, MDMR allows for systematic quantification of connectome reorganization across the whole-brain network without any prior parameters, and achieves superior sensitivity to NBS in distinguishing neurological disorders (Yang et al., 2021). Overall, these CWAS methods have exhibited excellent statistical power to locate aberrant brain structures or functional connectivity underlying neuropsychological disorders (Lee et al., 2020; Lin, Ni, Tseng, & Gau, 2020; Sheng et al., 2022; Yang et al., 2016). Nevertheless, the traditional statistical tools or machine learning implemented in these approaches are often deficient in modeling high order nonlinear intrinsic attributes and are limited in their performances. In addition, linking individual connectome variations to brain disorders with heterogeneous manifestations remains a significant challenge in CWAS implementation.

Deep learning is a different strategy for associating brain disorders with complex brain network variations. For instance, embedding connectome features into a low-dimensional space while preserving high-order nonlinear information, deep neural networks, such as graph neural networks (GNN), enable end-to-end disease identification at the individual level with promising performances. With the advantage of capturing the structural information contained within feature interactions in the graph domain, GNN has recently gained tremendous interests in the neuroscience field. One straightforward GNN architecture is to take each brain region as a graph node and iteratively aggregate high-order connectivity message among regions (denoted as the individual-graph architecture in this paper) (Kong et al., 2021; Ktena et al., 2018; Li et al., 2019; Zhang, Tetrel, Thirion, & Bellec, 2021; Zhao, Duka, Xie, Oathes, Calhoun et al., 2022). Another type of architecture represents a population of participants as one graph (denoted as population graph architecture in this paper), where each node is encoded with individual connectome features and each edge is manifested by inter-participant phenotypic information (Huang et al., 2020; Parisot et al., 2018a). Both architectures can predict individual neuropsychological disorders with performance similar to or surpassing that of convolutional neural network (CNN) (Jiang, Cao, Xu, Yang, & Zaiane, 2020; Li et al., 2021; Song et al., 2021).

However, existing GNN models potentially ignore local specificity in the brain. One of the core challenges is to precisely represent the complex brain connectome using a graph structure. Brain connectome features are previously considered translation-invariant by conventional graph methods; however, it has been criticized as too simplified to input whole brain connectome features into model training or embed them as identical (Li et al., 2021, 2021). Evidences suggest that connectome spatial locality plays a key role in characterizing neuropsychological patients into clusters while simultaneously locating disease-specific neuroimaging markers (Li, Satterthwaite, & Fan, 2017; Li, van Tol et al., 2014). Given that functional disrupted parcels are not spatially uniform distributed within the brain connectome, conventional graph learning approaches (e.g. GCN and GAT) are incapable to fully model brain connectome reorganizations in neuropsychological disorders. Unfortunately, few studies have addressed these issues.

In this regard, we propose a multivariate distance-based connectome network (MDCN) to associate region-specific and translation-variant connectome representations with brain disorders. The approach is inspired by the state-of-the-art CWAS approach, MDMR, where the classical brain connectome multivariate statistical module (Shehzad et al., 2014) is developed by the end-to-end GNN framework by extending the population distance matrix into a graph-learning framework. To tackle the local specificity issue, we propose to build locality-specific hyper-population graphs to characterize populations into clusters by assessing similarities among regional connectomes. In this regard, subtypes of neuropsychological diseases with high intra-class variations are preferable to be distinguished. Moreover, the network incorporating gradient and class activation maps facilitates model interpretability and relate connectivity-disease relationships instead of the conventional ANOVA-like analytic framework. We evaluated our novel method on two multi-center imaging databases: an Autism Spectrum Disorder (ASD) dataset, Autism Brain Imaging Data Exchange (ABIDE), consisting of 1022 participants from 17 centers, and an attention deficit hyperactivity disorder (ADHD)-200 global competition dataset, ADHD-200, with 572 participants from four sites for distinguishing ADHD. The effectiveness and robustness across various brain atlases and distance metrics were tested to validate whether the proposed computing framework could achieve both promising classification and interpretability. Extensive experiments demonstrated that our proposed MDCN provides new insights into connectome-wide association study. In summary, our contributions are:

- We established a multivariate distance-based connectome network (MDCN) to investigate the connectome reorganizations associated with heterogeneous neuropsychological disorders (i.e. ASD and ADHD) in a hyper-population graphs. The derived multiple graphs contributing to hyper-edges among individuals improve the connectivity representations compared to those of the homogeneous relations in a single graph.
- To address the spatial specificity in neuropsychological diseases, we proposed to model inter-parcel connectome representations by parcellation-wise attention and convolution modules. Moreover, parcellation-wise gradient class activation maps were investigated for relating key signatures. Both clinical explainability and high disease discriminative ability were achieved.
- Our MDCN identified brain connectome reorganization associated with ASD and ADHD. Similar parcel-specific connectome patterns suggest that the two disorders might fall on the same disease continuum and are likely to share overlapping neuropathological mechanisms.

## 2. Methodology

In the following sections, we first introduce the multivariate distance graph construction in Section 2.1. The details of the proposed MDCN are further investigated in Section 2.2. In addition, a method, termed parcellation-wise gradient and class activation map (p-GradCAM) to be used for interpretation is illustrated in Section 2.3. Finally, the algorithm scratch of our method with optimization is described in Section 2.4.

### 2.1. Multivariate distance graph construction

As shown in Fig. 1B, we calculated the distance between the connectivity patterns for every possible pairing of participants in a dataset, for each brain parcellation. The multivariate distance graph summarizes the individual differences that incorporate connectome features. The result was an $N \times N$ matrix of distances among $N$ participants for each region of interest. For instance, the 1st row or column of the matrix represents the (dis-) similarity between the 1st participant's whole-brain connectivity map and that of all the other participants'.

Assuming that the atlas is parceled into $M$ regions, the operation is carried out on each parcellation and repeated $M$ times. And totally a sequence of distance matrices $A^S$ is obtained as $A^S = (A^1, A^2, \ldots, A^M) \in R^{N \times N}$, where $A^i$ denotes the distance matrix built by the connectome features with respect to ROI $i$, $1 \leq i \leq M$. The distance matrix is obtained as follows, where

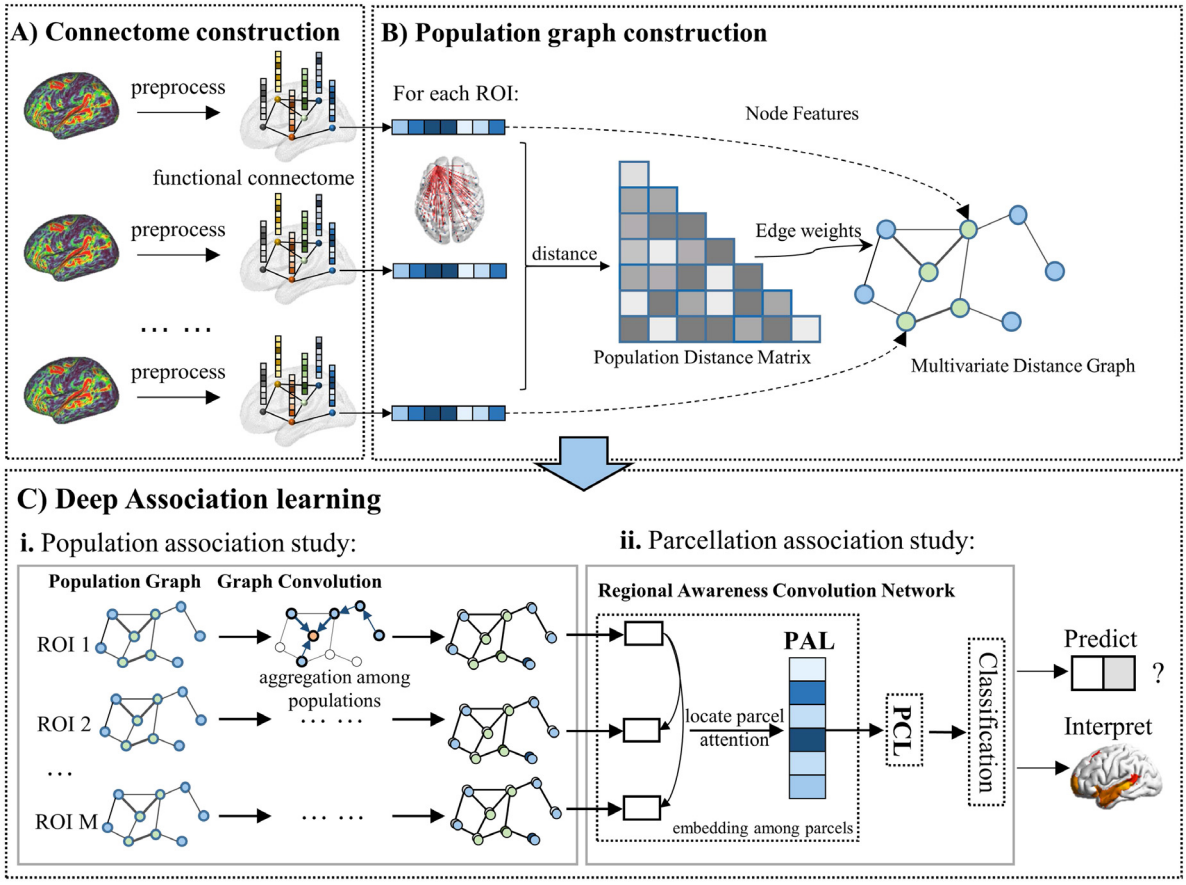


**Fig. 1.** The schematic representation of our proposed multivariate distance-based connectome network (MDCN) framework. (A) We first preprocessed the resting-state functional MR images and built the functional brain connectome, where each region of interest (ROI) is represented by a vector of connectivity connected to the region. (B) The functional connectivity derived from the processed functional MR images construct the multivariate distance graph incorporating the distance matrix. This step is repeated for each parcel and finally a sequence of graphs in the number of $\boldsymbol{M}$ is obtained. (C) The obtained graphs are fed into the framework for population and parcellation association learning by GCN and regional awareness convolution respectively. The parcellation association learning incorporates the Parcellation-wise Attention Layer (PAL) and Parcellation-wise Convolution Layer (PCL) modules for locating interests for each region.

the element $a^i_{u,v}$ represents the similarity between two nodes $V^i_u$, $V^i_v$ by calculating the distance measurement of the corresponding connectome features $X^i_u, X^i_v$:

$$a^i_{u,v} = dis\left(X^i_u, X^i_v\right) \cdot [1 + \sum_F \gamma_F(F_u, F_v)]$$

$$dis\left(X^i_u, X^i_v\right) = \exp(-\frac{[\rho(X^i_u, X^i_v)]^2}{2\sigma^2})$$

Here, a Gaussian probability density function was applied to formulate the distance in a Gaussian distribution with $\sigma$. The $\rho$ denotes the correlation distance and the other forms of distance are discussed further in the experiments. In addition, the distance is determined based on the similarities of the phenotype values, where $F$ denotes the covariance. Age, sex and site ID were included in our study. Function $\gamma$ is formulated as follows:

$$\gamma_F\left(F_u, F_v\right) = \begin{cases} 1 & if\ |F_u - F_v| < F_t \\ 0 & otherwise \end{cases}$$

In particular, $F_t$ represents the threshold for the corresponding equipment type. $F_t$ was set as 2 for age and 1 for sex and site ID. Subsequently, we combine the regional feature vector $X^S = (X^1, X^2, \ldots, X^M)$ and the adjacent matrix vector $A^S = (A^1, A^2, \ldots, A^M)$ to form multiple graphs $G^S = \left(G^1, G^2, \ldots, G^M\right)$.

### 2.2. Population-wise and parcellation-wise learning

A scratch of the MDCN framework is shown in Fig. 1C including a population association learning module and parcellation association learning module. To tackle the local specificity problem, we build a subgraph for each parcellation, where each subgraph is feasible to assess population similarities among regional connectome features. Considering the numerous multivariate distance graphs across large aggregated populations, we optimize the training process into two stages: population association learning and parcellation association learning.

**Population association learning.** For each subgraph, we implement the graph convolution for capturing the correlation between the participants and characterize them into clusters. In detail, spectral graph theory extends the convolution operation to graph structure data. The graph structure is represented by a Laplacian matrix in spectral graph analysis, where the Laplacian matrix and eigenvalues reflect the graph's topology properties. The Laplacian matrix is defined as $L = D - A$, where $L \in R^{N\times N}$ is the Laplacian matrix, $A \in R^{N\times N}$ is the adjacency matrix, $D$ is the degree matrix of the graph, and $N$ is the number of vertices. The eigenvalue matrix was obtained by decomposing the Laplacian matrix with $L = U\Lambda U^T$, where $U$ is the Fourier basis and $\Lambda \in R^{N\times N}$ is a diagonal matrix. For each vertex $i$, the node feature is $x_i \in R$. Subsequently, the graph Fourier transform and inverse Fourier

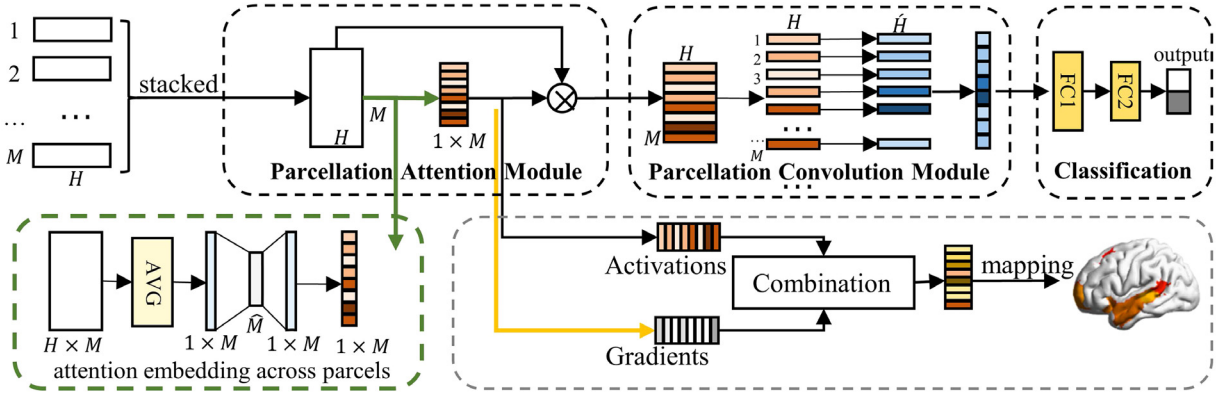


**Fig. 2.** The details of the regional awareness convolution network incorporating the p-GradCAM method. The backward gradients incorporating attention activations are provided to facilitate interpretability. AVG: an average pool operation for aggregate features from $\boldsymbol{H} \times \boldsymbol{M}$ into $\mathbf{1} \times \boldsymbol{M}$.

transform are defined as $\bar{x}_i = U^T x_i$ and $Ux_i = U\bar{x}_i$, respectively. The graph convolution of signals on graph is defined as $g_\theta \star_G x_i = Ug_\theta U^T x_i$, where $g$ denotes a graph convolution kernel and $\star_G$ denotes a graph convolution operation on graph $G$. Moreover, to simplify the calculation of the Laplace matrix, the $K$-order Chebyshev polynomials are adopted, which is defined as

$$g_\theta(\Lambda) \approx \sum_{k=0}^{K} \theta_k T_k(\tilde{\Lambda})$$

where $\tilde{\Lambda} = \frac{2}{\lambda_{max}}\Lambda - I_N$, where $I_N$ is the identity matrix, and $\lambda_{max}$ is the maximum eigenvalues of $\Lambda$. The $\theta \in R$ is a vector of polynomial coefficients. The Chebyshev polynomial is defined as $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ recursively, where $T_0(x) = 1$ and $T_1(x) = x$. With the $K$-order Chebyshev polynomials, the message from 1 to $K$-hop neighbors is aggregated to the center node.

**Parcellation association learning**. After embedding a regional connectome network into high order representations, the encoded representations are stacked into a sequence as $F \in R^{M \times H}$, which is a compact representation of the whole brain network for each sample. Conventional sequence models such as recurrent neural network (RNN) (Mikolov, Karafiát, Burget, Cernocký, & Khudanpur, 2010), long-short term network (LSTM) (Sak, Senior, & Beaufays, 2014), and gated recurrent unit (GRU) (Chung, Gulcehre, Cho, & Bengio, 2014), can be applied for embedding spatial neighborhood contextual information. However, these methods are limited in capturing sequential representations in order and have deficiency in relating to global dependencies. In this regard, we propose a regional awareness convolution network that leverages a parcellation attention module (PAM) to perform feature recalibration and adaptively concentrate on parcellations of interest.

As shown in Fig. 2, PAM aggregates the hidden features into a vector of size $M$ to produce a parcellation descriptor. The descriptor embeds the global distribution of the parcellation-wise feature responses. Parcellation-specific activations are then fed into an excitation operation to govern the excitation of each parcellation. The PAM is stacked in the first layer and generates a parcellation awareness embedded features. PAM can be obtained as follows:

$$\tilde{x} = x \cdot \sigma(W_{a2} \cdot \delta(W_{a1} \frac{1}{M} \sum_{i=1}^{M} F^i))$$

where $\sigma$ and $\delta$ are the sigmoid and ReLU function respectively. A bottleneck built by two fully connected layers with learnable weights $W_{a1}$ and $W_{a2}$ reduces and increases the embedding dimension. The final output was obtained by rescaling the transformation output with activations.

Moreover, a parcellation-wise convolution module (PCM) layer is leveraged to aggregate the spatial locality and relate topological relationships within an ROI. A parcellation-wise convolution filter is similar to a standard convolution layer with a kernel size of $H$, and locally filter features within a region of interest. PCM computes a weighted sum of features within an ROI, such as a convolution aggregating the local information of a patch. To improve the diversity of the embedding, multiple feature maps are obtained using different distinct convolutions.

### 2.3. Parcellation-wise gradient class activation maps

The lack of interpretability of deep learning models limits their clinical applications. Although the graph structural data lay a foundation for interpretable modeling, GNN still suffers from the opacity of information processing by mapping nodes and links into embedding (Liu, Feng, & Hu, 2022). To probe black-box neural network models, as well as for meaningful biomarker localization, we propose to provide parcellation-wise explanations by localizing key signatures with gradient and class activation maps. The activation maps are utilized to represent the importance as to which regions are responsible for the model's decision, while the gradient reflects the sensitivity changes. Related methods such as Grad-CAM++ (Chattopadhay, Sarkar, Howlader, & Balasubramanian, 2018) and Grad-CAM (Selvaraju et al., 2017), were originally proposed specifically for CNNs. In this study, we modified the setting by referring to Grad-CAM++ and sought insights in the attention module by considering the inner attention activation map and attention gradients.

In detail, given a stacked input $F_j$ for subject $j$, the regional awareness convolution network outputs the classification score $Y^c$, and the activation maps of the parcellation-wise attention module $S^c = \sigma(W_{a2} \cdot \delta(W_{a1} \frac{1}{M} \sum_{i=1}^{M} F_j^i))$ with $K$ feature maps, instead of averaging global gradients; the weighted average of the parcellation-wise gradients contributes to the overall decision. This is obtained as follows:

$$w_k^c = \sum_m \alpha_m^{kc} relu(\frac{\partial Y^c}{\partial S_m^k})$$

The $\alpha_m^{kc}$ denotes the weighting co-efficient for the gradient of the parcellation $m$ in terms of class $c$. By taking $Y^c = \sum_k w_k^c \cdot \sum_m S_m^k$ into consideration, the weights can be finally obtained with high-order derivatives, which are formulated as:

$$\alpha_k^c = \frac{\frac{\partial^2 Y^c}{(\partial S_m^k)^2}}{2\frac{\partial^2 Y^c}{(\partial S_m^k)^2} + \sum_m S_m^k \frac{\partial^3 Y^c}{(\partial S_m^k)^3}}$$

**Algorithm 1:** Pseudo-algorithm for MDCN

**Data:** Dataset $X$ and label y as $\{X_n, y_n | n = 1, ..., N\}$
**Input:** Split Dataset into 10-fold;
Draw training and Validation dataset input $X = \{X_n | n = 1, ..., N\}$;
Training Dataset Label $Y_{train} = \{Y_n | n = 1, ..., N_1\}$ ,$N_1 : N_2 = 9 : 1, N = N_1 + N_2$;
**Input:** Hyper-parameters: $dropout, n\_layers, hidden\_size$
**Output:** Optimized network parameters $W$;
Disease Prediction $\hat{Y}_{val} = \{\hat{Y}_n | n = 1, ..., N_2\}$; Explanation maps $\{R_n | n = 1, ..., N_2\}$
**for** $cv = 1, ..., 10$ **do**
  `// Stage 1`
  **for** $m = 1, ..., M$ **do**
    Calculate the distance matrix $A^m \in R^{N \times N}$ in terms of the brain region $m$;
    Generate the distance graph for region $G^m = \{X^m, A^m\}$;
    **for** $l = 1, ..., n\_layers$ **do**
      $H_1^m = X^m \cdot W_1$;
      $H_1^m \leftarrow g_\theta^m * H_1^m = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L}) H^m$;
      $H_1^m \leftarrow Drop(ReLu(H_1^m))$
    **end**
    $\hat{Y}_1^m \leftarrow Softmax(H_1^m)$;
    $L_1^m \leftarrow CrossEntropy(\hat{Y}_{1,train}^m), Y_{train})$
  **end**
  `// Stage 2`
  $F \leftarrow stack\{H_1^m | m = 1, ..., M\}$;
  $\widetilde{F} \leftarrow F \cdot (\sigma W_{a2} \cdot \delta(W_{a1} \frac{1}{M} \sum_{j=1}^{M} H_1^j))$;
  $H_2 \leftarrow PCM(\widetilde{F})$;
  $\hat{Y}_2 \leftarrow ReadOut(H_2)$;
  $L_2 \leftarrow CrossEntropy(\hat{Y}_{2,train}^m), Y_{train})$;
  `// Predict and Output p-GradCAM`
  Predict: $\hat{Y}_{val} \leftarrow \hat{Y}_{2,val}$;
  Interpret: Generate individual maps: $R = relu(\sum_k w_k^c \cdot S^{km})$
**end**

The final maps $R$ are obtained based on the weighted combination of the feature maps:

$$R = relu\left(\sum_{k=1}^{K} w_k^c \cdot S_m^k\right)$$

### 2.4. Optimization

Considering the increasing parameters of $M$ graph networks, it is difficult to optimize the framework simultaneously. The supervised learning in our framework is decoupled into two stages: optimization of the graph convolution network and the regional awareness convolution network.

In both stages, the embedded representations of the graph and convolution networks are averaged and classified by a two-layer multiple perception followed by a *softmax* function:

$$\hat{y}_2 = softmax\left((H_2 W_{21} + b_{21}) W_{22} + b_{22}\right)$$

where $W$ and $b$ are learnable parameters of a two-layer multiple layer perceptions (MLP), as the readout function for embedded connectome features. The loss function is formulated as a cross-entropy function by learning from:

$$J = -\sum_n [y^{(n)} \log\left(\hat{y}^n\right) + \left(1 - y^{(n)}\right) \log\left(1 - \hat{y}^n\right)]$$

A pseudo-algorithm is presented in Algorithm **1**:

## 3. Experiments

### 3.1. Datasets

We examined resting-state functional MR images (fMRI) using two largely aggregated multicenter datasets. (1) The first dataset was the ABIDE-I database, where T1 structural brain images, resting-state functional MR images, and phenotypic information from 17 different imaging sites were aggregated for each patient. The initial ABIDE-I included 505 individuals diagnosed with ASD and 530 controls. A total of 502 sex-matched participants with ASD and 520 HC were included in the evaluation in our study. (2) The second dataset was the ADHD-200 database. The ADHD-200 dataset was first publicly released during an ADHD-200 global competition. The training set of the competition was included as a cohort including 488 healthy controls and 280 patients meeting the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) criteria for ADHD. Age and sex-matched participants (311 HC and 261 ADHD) from four centers were included. Table 1 summarizes the key demographic details and statistical results.

### 3.2. Preprocessing

**The ABIDE-I dataset** All images were preprocessed using the configurable pipeline for the analysis of connectomes (C-PAC),

**Table 1**
Statistical results on demographical information on two datasets. M and F denote the numbers of the male and female. The average and standard deviation of age are displayed (mean ± std).

| | | Normal | Patient | F-statistic | ANOVA $p$ |
|---|---|---|---|---|---|
| ABIDE | Sex (M/F) | 445/75 | 443/59 | 1.264 | 0.206 |
| | Age (years) | 17.21 (7.93) | 17.06 (8.34) | 0.303 | 0.762 |
| ADHD-200 | Sex (M/F) | 230/81 | 209/52 | 1.728 | 0.085 |
| | Age (years) | 11.50 (2.37) | 11.18 (2.42) | −1.599 | 0.112 |

including skull striping, slice timing correction, motion correction, global mean intensity normalization, nuisance signal regression with 24 motion parameters, and band-pass filtering (0.01–0.08 Hz). Finally, the functional images were registered into the standard anatomical space (MNI152) to allow cross-participant comparisons. The nuisance variable regression was modeled using 24 motion parameters. The ABIDE dataset can be obtained online.[2]

**The ADHD-200 dataset:** The images were preprocessed by the Athena[3] pipeline incorporating AFNI (Cox, 1996) and FSL (Smith et al., 2004) neuroimaging tools. The preprocessed images are publicly available and can be obtained online.[4] The pipeline involves removing the first four volumes, slice timing correction, realignment to correct for motion, and linear transformation between the mean functional volume and the corresponding structural MRI. The functional images were transformed into MNI-152 space incorporating T1-weighted MRI into the MNI nonlinear warp. Nuisance regression models were used to remove noise and head drifts in the time series. The denoised time series was then band-pass filtered (0.009 Hz–0.08 Hz).

The mean time series for a set of regions extracted from the Schaefer template (Schaefer et al., 2018) were computed and normalized to zero mean and unit variance. Pearson's Correlation Coefficient was used to measure functional connectivity. Regional time series were obtained by extracting the Schaefer atlas, which was parceled by a gradient-weighted Markov random field approach that identifies cortical parcels ranging from 100 to 1000 parcels. In our implementation, we employed the Schaefer atlas with 100 parcels, termed Schaefer-100, for the main evaluation, and tested classification performances and biological explanations with Schaefer-200 and Schaefer-400 at different resolutions. Each parcel was matched to a corresponding network in the seven network parcellations by (Yeo et al., 2011) including the default mode (DMN), visual, the somatomotor, dorsal attention, salience/ventral attention, limbic, and control networks.

### 3.3. Implementations

Our experiments were carried out on an NVIDIA Tesla V100 with 32 GB memory. In the graph construction, the threshold $F_t$ was set as 2 for age, and 1 for sex and site ID respectively. The $\sigma$ in a Gaussian distribution is set as 2 and the ROI number in this study is 100. A two-layer multi-layer perception followed by a softmax layer is implemented as the classifier in MDCN. For the MDCN, the $K$ order of Chebyshev Polynomials is set as 3. The GCN parameters including hidden size and number of layers are optimized using a grid search, where the hidden size is in the range of [8, 16, 32, 64], and graph convolution layers in [1, 2, 3]. A dropout is included to improve the generalization ability with a drop rate of 0.3. The learning rate within the network is set as 0.005, and the weight decay for regularization is 5e−4.

For a fair comparison with other state of the art methods on the ABIDE/ADHD-200 dataset, we trained and tested our model using 10-fold stratified cross-validation. The 10-fold cross-validation facilitates the evaluation and comparison with the state-of-the-art, where the strategy is adopted in most related studies. Moreover, to evaluate the performance of comparative methods, prediction accuracy (ACC), sensitivity (SEN), specificity (SPE), F1 score (F1), and area under the curve (AUC) are used for evaluation. Our MDCN is trained in a two-stage manner. In the first stage, the representations are obtained for the second stage of training when the validation loss reaches the lowest in 500 epochs. The second stage fits the embedded features of the training dataset and validation dataset and is trained in 500 epochs.

### 3.4. Competitive methods

To test the MCDN method, several CWAS-based methods and deep learning models developed for the brain connectome learning were evaluated. These methods include (1) SVM: the functional connectivity (FC) matrix with the upper triangle was flattened to a vector form with $100 \times 99/2 = 4950$ features. A support vector machine (SVM) is trained for classification. SVM is considered a robust approach for classification and could also be tested as a baseline for performance improvement comparison. (2) NBS/MDMR-SVM (Shehzad et al., 2014; Zalesky et al., 2010): The NBS and MDMR algorithms are two state-of-the-art CWAS methods for revealing the widespread brain network reorganization of diseases. In the experiments, numerous functional connectivity features were filtered by NBS and MDMR to locate and distinguish key connectome features. Given that the NBS parameter threshold is difficult to choose, we determined this value by a grid search in the range of 0.05 to 0.15 with steps of 0.005, resulting in features with different sizes for NBS. In addition, key regions with significant differences (p-value $< 0.05$) were detected by MDMR, and connections among these regions were selected as input features for MDMR. Finally, a support vector machine was trained to evaluate the selected features. (3) Graph convolutional network (GCN) (Parisot et al., 2018a): We employed a semi-supervised GCN with features selected via recursive feature elimination (RFE). Each node represents a subject, and the feature size of each node is reduced to 1000 by RFE. Population similarities measured by age and site ID phenotypes were considered as the edge weight. The parameters were set according to a previous study (Parisot et al., 2018a). (4) BrainNetCNN (Kawahara et al., 2017): BrainNetCNN is proposed to train the connectome matrix using convolution based on edge-to-edge (E2E), edge-to-node (E2N), and node-to-graph (N2G) layers. Two E2E layers with a channel size of 32, an E2N layer with 64 output features, and an N2G layer with 30 output features were used for training. The drop rate is set as 0.5 by referring to the original architecture. (5) Brain Graph Neural Network (BrainGNN) (Li et al., 2021): The BrainGNN proposed ROI-aware graph convolutional layers and ROI-selection pooling layers for neurological biomarker prediction at the group and individual levels, thus outperforming a series of fMRI image analysis methods. We implemented the original network architecture and modify the parameters by setting $K^{(0)} = K^{(1)} = 16, d^{(0)} = 100, d^{(1)} = d^{(2)} = 32$, and tuning the $\lambda_1$ and $\lambda_2$ in a grid search of (0, 1) with a step of 0.1. (6) Hypergraph Neural Network (HGNN) (Feng, You, Zhang, Ji, & Gao, 2019) and Dynamic Hypergraph Neural Network (DHGNN) (Jiang, Wei, Feng, Cao, & Gao, 2019): the proposed hyperedges are feasible to capture distinctive associations among parcellations. We follow the original architecture

[2] https://fcon_1000.projects.nitrc.org/indi/abide/.

[3] http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline.

[4] http://neurobureau.projects.nitrc.org/ADHD200/Introduction.html.

**Table 2**
Comparisons of the classification performance (mean ± standard deviation) on ABIDE and ADHD-200 datasets with different models. The best result for each row is shown in bold.

| Dataset | ABIDE | | | | | ADHD-200 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ACC | SEN | SPE | F1 | AUC | ACC | SEN | SPE | F1 | AUC |
| SVM | 65.56 ± 5.78 | 70.00 ± 7.86 | 60.99 ± 11.28 | 67.22 ± 6.82 | 71.42 ± 7.86 | 54.02 ± 6.98 | 36.08 ± 24.64 | 69.71 ± 15.61 | 37.55 ± 21.68 | 49.64 ± 15.57 |
| NBS-SVM | 63.51 ± 3.60 | 72.85 ± 6.44 | 53.80 ± 9.16 | 67.25 ± 3.15 | 67.04 ± 3.77 | 58.73 ± 7.46 | 49.34 ± 23.27 | 66.91 ± 13.54 | 49.61 ± 18.56 | 60.95 ± 11.55 |
| MDMR-SVM | 67.43 ± 5.14 | 72.11 ± 8.03 | 62.59 ± 11.87 | 69.20 ± 4.55 | 73.17 ± 6.18 | 56.20 ± 5.33 | 42.89 ± 24.74 | 67.85 ± 14.45 | 43.59 ± 20.12 | 55.64 ± 12.56 |
| GCN | 69.11 ± 3.21 | 67.09 ± 4.27 | 71.22 ± 3.17 | 66.37 ± 5.48 | 72.56 ± 3.21 | 61.02 ± 3.64 | 46.48 ± 9.34 | 70.12 ± 9.78 | 51.26 ± 5.40 | 55.78 ± 7.35 |
| BrainGNN | 69.31 ± 3.03 | 68.18 ± 11.05 | 68.64 ± 13.77 | 69.08 ± 5.77 | 69.52 ± 3.52 | 66.63 ± 5.41 | 70.10 ± 12.77 | 63.97 ± 10.79 | 66.04 ± 7.92 | 64.03 ± 3.88 |
| BrainNetCNN | 69.73 ± 2.76 | 66.73 ± 9.93 | 72.54 ± 8.26 | 67.73 ± 5.09 | 71.99 ± 3.06 | 63.77 ± 3.84 | 69.87 ± 12.76 | 58.37 ± 13.57 | 63.95 ± 4.83 | 62.97 ± 5.10 |
| HGNN | 70.96 ± 4.73 | 69.47 ± 8.06 | **77.25 ± 9.44** | 58.40 ± 16.12 | 69.37 ± 6.80 | 65.39 ± 4.22 | 64.63 ± 4.37 | 71.12 ± 12.92 | 59.12 ± 11.98 | 62.94 ± 4.70 |
| DHGNN | 71.45 ± 5.82 | 71.86 ± 16.89 | 71.20 ± 7.57 | 67.32 ± 15.93 | 72.39 ± 6.87 | 66.49 ± 5.27 | 64.35 ± 6.13 | 66.00 ± 14.48 | 60.32 ± 9.78 | 65.39 ± 5.33 |
| HI-GCN | 69.31 ± 5.14 | 73.00 ± 10.63 | 70.26 ± 4.57 | 71.95 ± 6.16 | 69.21 ± 5.14 | 64.23 ± 4.39 | 63.13 ± 11.94 | **66.19 ± 8.99** | 55.62 ± 9.74 | 58.97 ± 4.45 |
| TE-HI-GCN | 71.08 ± 5.41 | 72.38 ± 10.59 | 71.92 ± 8.87 | 71.24 ± 4.71 | 71.04 ± 5.53 | 66.55 ± 5.30 | 60.43 ± 15.02 | 62.64 ± 19.81 | 58.70 ± 15.31 | 61.54 ± 7.26 |
| MDCN | **72.41 ± 2.51** | **73.16 ± 8.82** | 71.73 ± 6.55 | **72.01 ± 4.12** | **73.91 ± 2.76** | **67.45 ± 5.32** | **71.97 ± 11.22** | 62.39 ± 11.24 | **66.86 ± 6.91** | **65.39 ± 6.04** |

and modify the classifiers for graph classification task. The hidden size within [16, 32, 64, 128]. (7) Hierarchical GCN (HI-GCN) (Jiang et al., 2020) and Ensemble of Transfer Hierarchical Graph Convolutional Networks (TE-HI-GCN) (Li et al., 2022): The hierarchical architecture is proposed to associate graph topology information with the participants' similarities. And multiple thresholds of connectivity are implemented to build graph kernels. For a fare comparison with other networks, the transfer learning part is discarded. The original architecture is used and the ROI number is set as 100.

## 4. Results

### 4.1. Evaluation of disease prediction

We evaluate the disease prediction performances in terms of accuracy, sensitivity, specificity, F1-score and area under the curve (AUC). As shown in Table 2, the proposed method surpasses the other related methods (72.41% ACC and 73.91% AUC for ASD classification, and 67.45% ACC and 65.39% AUC for ADHD classification). Deep learning methods, such as BrainNetCNN and BrainGNN, outperform CWAS-based methods in terms of classification performance, indicating that the embedded underlying nonlinear interactions and dependencies can be used to improve the representations and play a key role in diagnosis. Moreover, in the classification of ASD, the MDMR-based model outperforms the NBS-based model (MDMR-SVM: 67.43% ACC; NBS-SVM: 63.51% ACC), while the opposite is true in the classification of ADHD (MDMR-SVM: 56.20% ACC; NBS-SVM: 58.73% ACC). The same is true for BrainNetCNN and BrainGNN, indicating that these methods are limited in their robustness and generalization ability across different tasks. Moreover, HGNN and DHGNN achieves better performances than BrainNetCNN and BrainGNN. This is because the hyperedge mechanism could improve the learning of associations among parcellations by distinctive measurements. Compared with hyperedge, HI-GCN and TE-HI-GCN builds hierarchical structures to capture associations among parcelations, which is similar to the hyperedge mechanism. Accordingly, these approaches achieves comparable performances in most cases. Compared with these methods, MDCN achieves the best on both two datasets.

In addition, the consistent improvements in both datasets of our MDCN demonstrate the robustness of the model. These improvements may result from three causes. First, the functional connectome in high complexity is embedded with deep, complex, and nonlinear deep learning models. Convolutions are particularly useful when the features are too complex to be designed. Moreover, our model is implemented based on graph theory and has the flexibility to capture the local and global graph topological attributes. Finally, our method builds associations among populations and parcellations across the whole brain without the burden of feature selection.

### 4.2. Evaluation of model interpretability

The ability of deep learning methods to model nonlinear representations is vital but can make interpretation challenging (Yan et al., 2022). Apart from other deep learning methods, our proposed MDCN incorporating p-GradCAM is a CWAS-guided framework that is capable of performing its interpretability for both individual and group differences.

#### 4.2.1. Individual maps derived from p-GradCAM

Fig. 3 visualizes the gradient and class activation maps on two datasets, where three participants of each subgroup were randomly sampled for comparison. The p-GradCAM maps were aligned to the anatomical space using the derived values. The results of the two datasets are displayed from top (ABIDE) to bottom (ADHD-200), where red shows the amplitude of the map values; for example, dark red indicates that the regions tend to contribute to the functional connectome and are more correlated. The maps of healthy participants are shown with relatively consistent spatial distributions, whereas those of patients with ASD/ADHD are considered distinctive. This conforms to the heterogeneity of psychological diseases in the clinic, which results in inconsistent disorganization in functional changes. As shown in Fig. 3A, compared with patients with ASD, healthy participants tended to achieve higher values in the medial and dorsal prefrontal cortex regions. This indicates that functional changes may exist in these patients. The same phenomenon also occurs in the ADHD-200 dataset, as shown in Fig. 3B, where patients with ADHD show lower amplitudes in the prefrontal cortex. Moreover, we can observe that the regions located in the hippocampus of HC and the patients are considered distinctive, where the amplitudes of regions of the medial and dorsal prefrontal cortex in the HC are much higher than those in ASD and ADHD.

#### 4.2.2. Group analysis

Notably, the regions of interest of HC are attentively located compared to those of patients. The class activation score is a relative score; thus, a region of high interest for HC might show a decrease in that of the patients. An independent t-test was performed for the group statistical evaluation. Bonferroni correction was applied to control the false-positive rate. A p-value $< 0.05$ after correction was determined significance within the experiments.

Fig. 4 summarizes the spatial distribution of the located patterns with significant differences (adjusted p-value $< 0.05$). For the ABIDE dataset, four brain seed regions were located. These regions belong to three functional subnetworks: default mode, limbic, and control networks. For the ADHD-200 dataset, the default mode network was found to be significantly different. Statistical results showed that brain connectivity patterns within these key regions were significantly decreased compared to those of healthy participants. The detailed results are presented in Table 3. The medial and dorsal prefrontal cortex was found with significant differences in both datasets.

**Table 3**
Brain regions with significant differences on ABIDE and ADHD-200 datasets.

| | Brain regions | T-Statistic value | p-value | Adjusted p-value |
|---|---|---|---|---|
| ABIDE | LH_Vis_1 | 3.978 | <0.001 | 0.007 |
| | LH_SalVentAttn_Med_2 | −3.770 | <0.001 | 0.017 |
| | LH_Default_PFC_3 | −3.964 | <0.001 | 0.008 |
| | RH_Default_PFCdPFCm_3 | −3.602 | <0.001 | 0.033 |
| ADHD-200 | RH_Default_PFCdPFCm_2 | −3.668 | <0.001 | 0.027 |

**Vis**: The Visual Network; **SalVentAtten**: The Salience and Ventral Attention Network; **Default:** The Default Mode Network; **Med**: Medial; **PFC**: prefrontal cortex; **PFCdPFCm**: dorsal and medial prefrontal cortex; **LH**: the left hand; **RH**: the right hand.

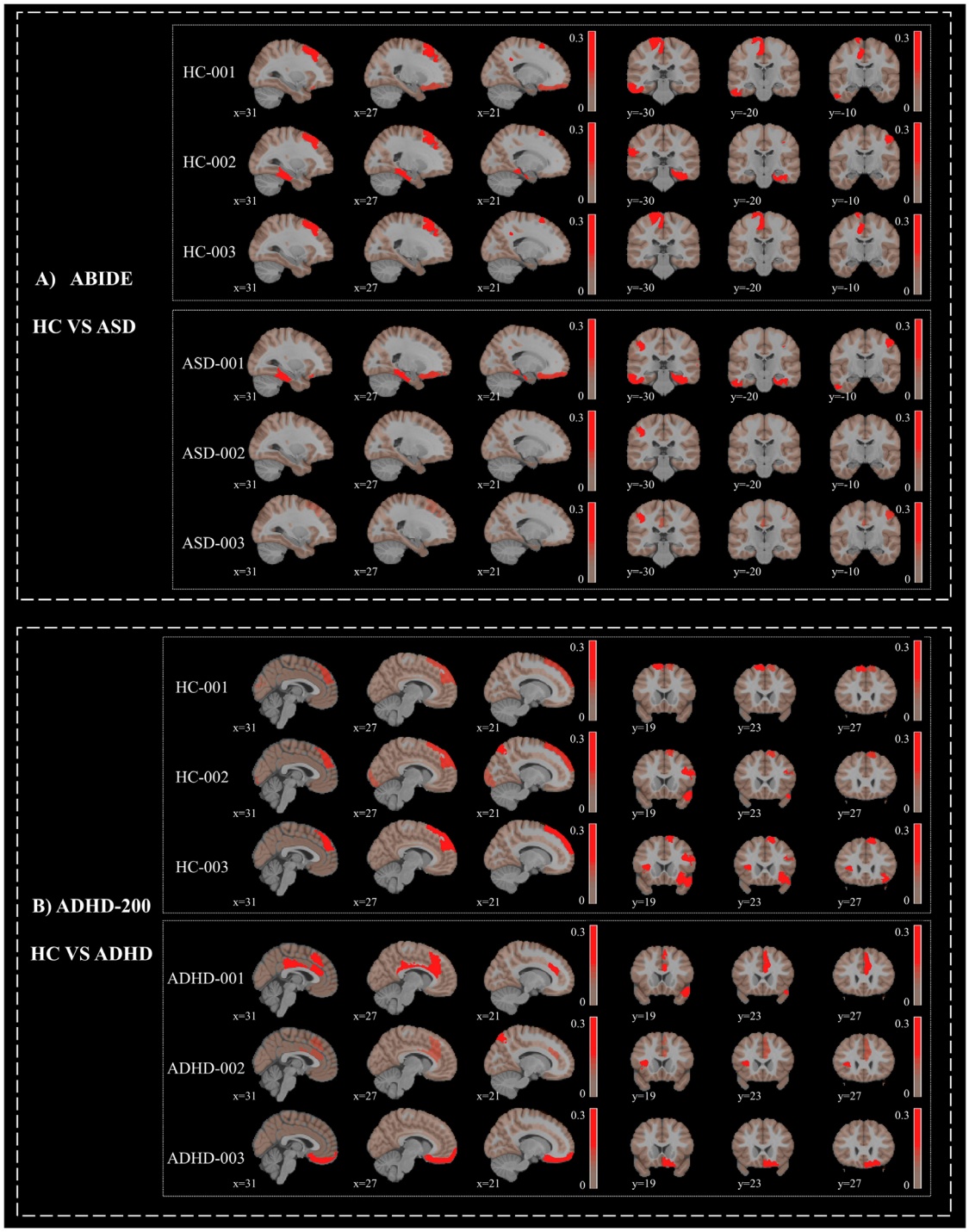


**Fig. 3.** Examples on p-GradCAM maps for three individuals of HC group and the patient group with ASD/ADHD respectively. A threshold of 0.3 is used for better visualization over all the maps. Three maps from axial, coronal panels are displayed from left to right respectively.

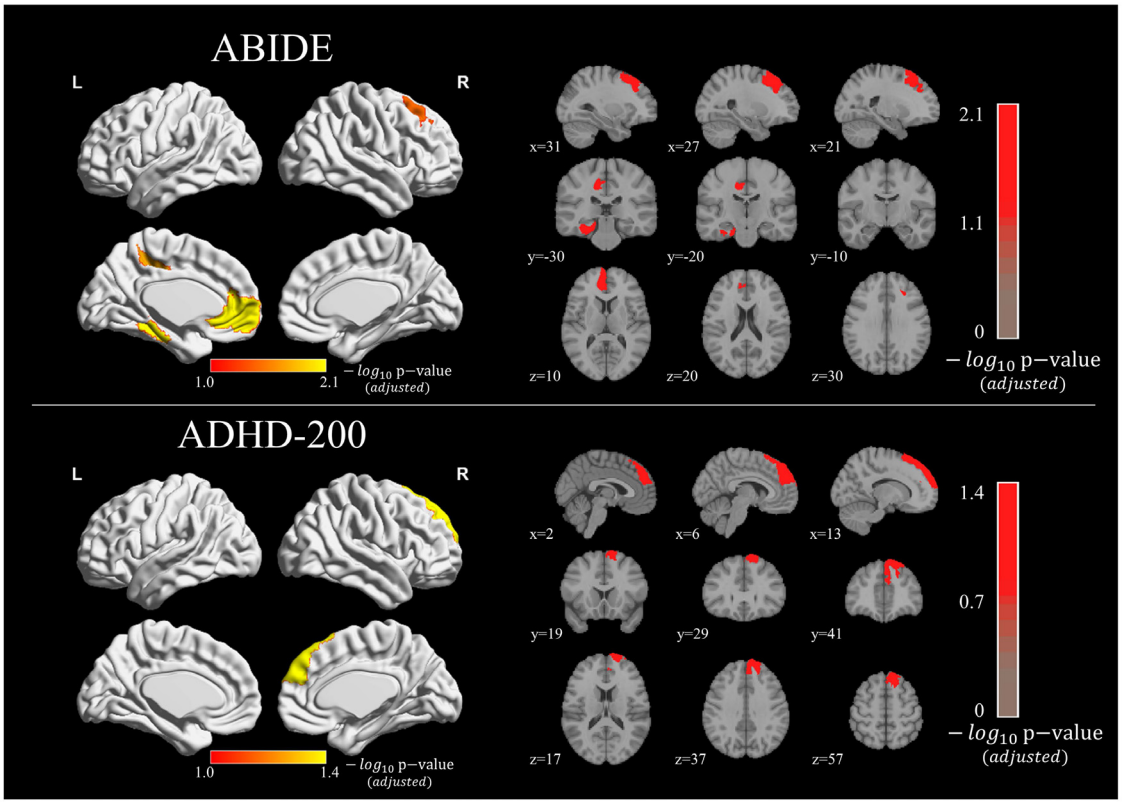


**Fig. 4.** The located significant patterns on ABIDE and ADHD-200 datasets are shown from top to bottom (adjusted p-value $< 0.05$). The surface plots (left) demonstrate the scratch of spatial distributed changes with significant difference. The details are further displayed in slices from axial, coronal, and sagittal panels (right). Significant differences were found in the medial and dorsal prefrontal cortex in both ABIDE and ADHD-200 datasets.

**Table 4**
Regions with significant differences of Schaefer-100, Schaefer-200, and Schaefer-400 atlases. Only regions with significant differences (adjusted p-value <0.05) are listed.

| | Brain regions | T-statistic value | p-value | Adjusted p-value |
|---|---|---|---|---|
| Schaefer-100 | LH_Vis_1 | 3.977 | <0.001 | 0.007 |
| | LH_SalVentAttn_Med_2 | −3.770 | <0.001 | 0.017 |
| | LH_Default_PFC_3 | −3.963 | <0.001 | 0.008 |
| | RH_Default_PFCdPFCm_3 | −3.601 | <0.001 | 0.033 |
| Schaefer-200 | RH_Default_PFCdPFCm_7 | −3.736 | <0.001 | 0.039 |
| Schaefer-400 | RH_Default_PFCdPFCm_9 | −3.890 | <0.001 | 0.043 |

**Vis**: The Visual Network; **SalVentAtten**: The Salience and Ventral Attention Network; **Default:** The Default Mode Network; **Med**: Medial; **PFC**: prefrontal cortex; **PFCdPFCm**: dorsal and medial prefrontal cortex; **LH**: the left hand; **RH**: the right hand.

#### 4.2.3. Comparison of MDCN findings with MDMR

One remarkable feature of our proposed MDCN is that it leverages the population distance matrix to measure inter- and intra-group similarities, which are also shared by MDMR. Accordingly, we compared the key biomarkers detected by MDCN and MDMR approaches. Similarly, Bonferroni correction was applied to control for the false-positive rate. After correction, a p-value $< 0.05$ after correction was determined significant.

As shown in Fig. 5, we detected overlapping regions in both diseases, especially in the dorsal and medial prefrontal cortex. In both datasets, the MDMR method locates regions mainly on the default mode network including regions in the temporal cortex, the prefrontal cortex, and the posterior cingulate cortex in the ABIDE dataset, and the prefrontal cortex in the ADHD-200 dataset. Specifically, the overlap on the dorsal and medial prefrontal cortex was located using both methods (MDMR p-value: <0.001; MDCN p-value: 0.033 in the region of RH_Default_PFCdPFCm_3). Similarly, an overlapping region was also detected in the ADHD-200 dataset, where the dorsal and medial prefrontal cortex was found (MDMR p-value: <0.001; MDCN p-value: 0.038 in the region of RH_Default_PFCdPFCm_2). These common findings demonstrate the feasibility of our proposed MDCN, which is implemented based on deep learning, while simultaneously pinpointing the key brain connectivity signature for ASD and ADHD.

### 4.3. Sensitivity analysis

To examine the effect of the atlas, we parcellated the brain at different spatial resolutions, including 100, 200, and 400 parcels based on the Schaefer atlas (i.e., Schaefer-100, Schaefer-200, and Schaefer-400). The experiments were repeated at each atlas for comparison. As the scale of regions decreased, only one region with a significant difference among the three atlases was found, where the connectivity in the dorsal and medial prefrontal cortex of ASD patients decreased compared to that of HCs, as demonstrated in Table 4. The findings in the dorsal and medial prefrontal cortex are consistent with that in Schaefer-100 (the right dorsal and medial PFC in Schaefer-100: adjusted p = 0.033; in Schaefer-200: adjusted p = 0.039; Schaefer-400: adjusted $p$ =

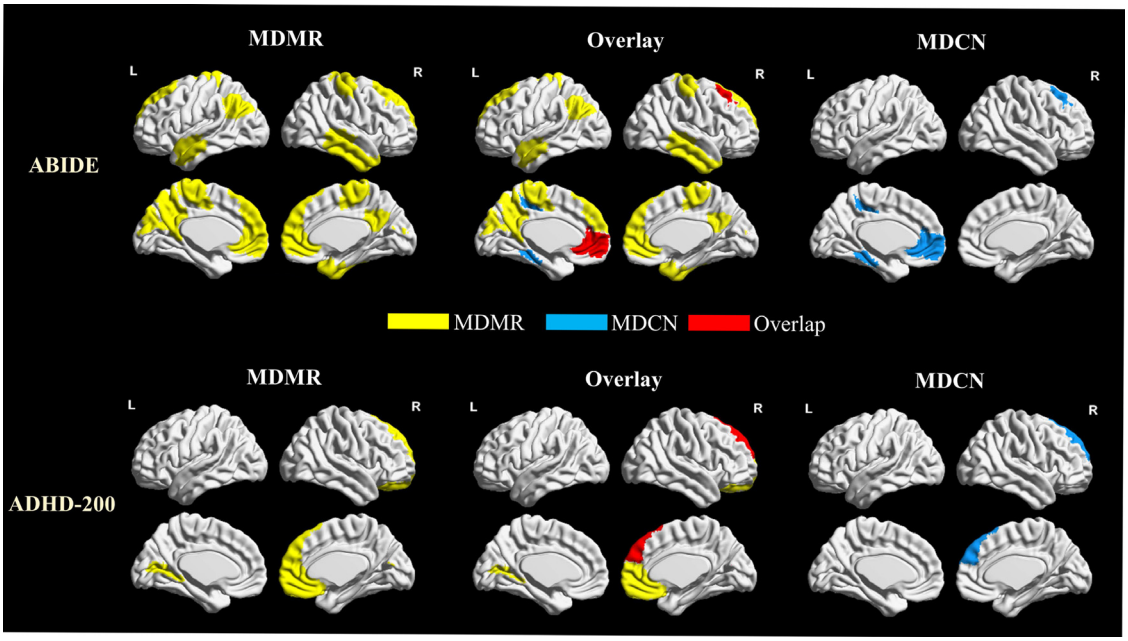


**Fig. 5.** Comparison of MDCN and MDMR for the connectivity analysis, where the evaluations on ABIDE and ADHD-200 datasets are displayed from top to bottom. The statistical maps (adjusted p-value $<0.05$) returned from the MDMR are shown in yellow, those from the MDCN are shown in blue, and the overlaps between the two are shown in red. The overlap on the dorsal and medial prefrontal cortex was located by MDCN and MDMR.

0.042). To summarize, we ascertained the spatial distribution in Fig. 6A. Although the specific regions of different atlases show slight spatial fluctuations, the consistent results in the dorsal and medial prefrontal cortex could indicate the power of the interpretability of our framework. In addition, as shown in Fig. 6B, the highest classification accuracy was achieved with the Schaefer-400 atlas, whereas the AUC was the best with the Schaefer-100 atlas, indicating that the generalization ability of the model using Schaefer-100 is the highest.

In addition, we examined the effect of distance measurements, including the Euclidean, Chebyshev, and cosine similarity distances. Fig. 6 (C) visualizes the statistical maps across various distance measures and the detailed values are listed in Table 5. Consistent results were found in the medial and dorsal prefrontal cortex (RH_Default_PFCdPFCm_3). In addition, a significant difference in the correlation and Chebyshev distance was found in the left prefrontal cortex (adjusted p-value with correlation distance: 0.033; with Chebyshev distance: 0.036). Moreover, Fig. 6D displays the classification effect across distance measures, where the best model was achieved by the correlation metric with the highest AUC (73.91%) and comparable accuracy (72.41%). Nevertheless, the faint fluctuations across atlases and distance metrics indicate that our MDCN is considered slightly robust and sensitive to the atlas and distance metrics.

## 5. Discussion

In this study, we propose a unified connectome-wide deep learning framework, MDCN, to interpret the associations between brain network reorganization and clinical phenotypes. MDCN can efficiently map individual brain network signatures by high-order nonlinear message passing among both populations and parcellations and further leverage this signature for disease identification in an end-to-end manner. Our results demonstrated that MDCN achieved the best performance in distinguishing ASD and ADHD from healthy controls based on resting-state fMRI data among other state-of-the-art CWAS-based approaches and deep learning networks. In particular, MDCN allows visualization of the disease-related connectome patterns on each individual by the highly overlapping regions observed between MDCN and MDMR on the two large datasets. The substantial correspondence and outstanding classification performance demonstrate the advantage of our unified connectome-wide learning framework over the two traditional constituent steps (i.e., connectome mapping and task learning).

### 5.1. The advantages of MDCN for brain connectome study

The proposed MDCN includes (1) graph convolution layers for population association studies, (2) PAM and PCM layers for parcellation association studies and (3) an interpretability module, p-GradCAM. This leads to several advantages for the MDCN.

Graph convolutions that act as low-pass filtering (Nt & Maehara, 2019) benefits from leveraging higher-order association information that could arise from population clusters with high similarity. Modeling the population associations in graphs maximizes the discrepancies of brain network features in heterogeneity as well as relating dependencies among neighbors. Moreover, the population graph was built based on features within each parcellation and repeated across the whole brain. The derived multiple graphs contribute to hyper-edges among individuals, which improves the connectivity representations compared to those of the homogeneous relations in a single graph. Compared with most existing graph methods that build linear relations in a graph, our method benefits from learning nonlinear associations between participants using the hyper-edges. Nevertheless, MDCN is feasible for embedding each region separately in population graphs. In this way, the regional disease-specific structural information contained in the brain connectome can be utilized. Finally, the PAM and PCM layers benefit from parcellation attention learning and are feasible for modeling region-specific connectome representations. Our proposed MDCN showed superior prediction accuracy and improves the accuracy by 2.41%, given that an accuracy of 70% might be an intrinsic limitation of the ABIDE dataset (Parisot et al., 2018b). Notably, our proposed MDCN benefits from a powerful graph learning architecture instead of conventional machine learning or statistical models. For comparison, we implemented the support vector machine with

**Table 5**
Regions of various distance measures with significant differences (adjusted p-value <0.05) are listed.

| | Brain regions | T-Statistic value | p-value | Adjusted p-value |
|---|---|---|---|---|
| Correlation | LH_Vis_1 | 3.977 | <0.001 | 0.007 |
| | LH_SalVentAttn_Med_2 | −3.770 | <0.001 | 0.017 |
| | LH_Default_PFC_3 | −3.963 | <0.001 | 0.008 |
| | RH_Default_PFCdPFCm_3 | −3.601 | <0.001 | 0.033 |
| Euclidean | LH_Default_PFC_7 | −3.684 | <0.001 | 0.024 |
| | RH_Default_PFCdPFCm_3 | −3.864 | <0.001 | 0.012 |
| Cosine | RH_Default_PFCdPFCm_3 | −4.084 | <0.001 | 0.005 |
| Chebyshev | LH_Default_PFC_3 | −3.946 | <0.001 | 0.008 |
| | RH_Vis_7 | −5.045 | <0.001 | <0.001 |
| | RH_SomMot_4 | 4.713 | <0.001 | <0.001 |
| | RH_Default_PFCdPFCm_3 | −3.580 | <0.001 | 0.036 |

**Vis**: The Visual Network; **SalVentAtten**: The Salience and Ventral Attention Network; **Default:** The Default Mode Network; **SomMot**: The Somatomotor Network; **Med**: Medial; **PFC**: prefrontal cortex; **PFCdPFCm**: dorsal and medial prefrontal cortex; **LH**: the left hand; **RH**: the right hand.

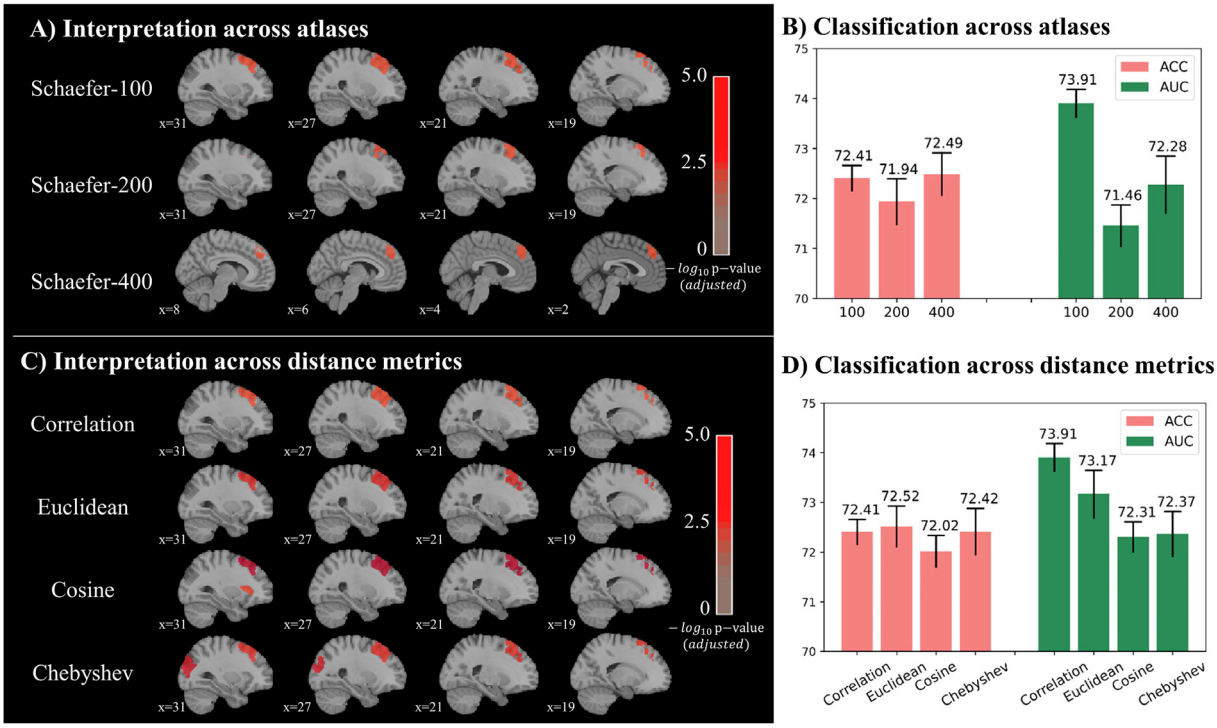


**Fig. 6.** Visualizations of located brain regional features and classification comparisons across multiple atlases and distance metrics. (A) The statistical maps across different atlases, where the sagittal views of Schaefer-100, Schaefer-200, and Schaefer-400 are displayed from top to bottom respectively. (C) The statistical maps across different distance metrics include the correlation, Euclidean distance, cosine similarity distance, and Chebyshev distance. The regions with significant differences are shown in red. The corresponding classification results are displayed in (B) and (D) respectively, where the classification accuracy (ACC) is shown in red and the area under the curve (AUC) in green.

statistical CWAS methods (NBS and MDMR) for feature engineering. These methods are fed with selected features filtered by statistical analysis across the whole brain but perform poorly in classification (MDMR-67.43% VS MDCN-72.41% for ASD; NBS-58.73% VS MDCN-67.45% for ADHD). This result suggests that graph learning is more efficient in capturing and embedding meaningful representative connectome features.

Although deep learning models are commonly criticized for black-box problems, several recent approaches have been proposed to improve their explanatory ability. Interpretable local surrogates (e.g. LIME Ribeiro, Singh, & Guestrin, 2016) were found to provide reliable explanations but were limited in local fidelity, where the explanations must be locally faithful to be meaningful (Katarya, Sharma, Soni, & Rath, 2022). Gradient-based methods, e.g. CAM (Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016) and GradCAM (Selvaraju et al., 2017), are most frequently used to derive interpretable visualization for disease identification. Our proposed p-GradCAM further takes advantage of the gradients of the parcellation-wise attention module instead of a global average pooling operation, which explicitly relates to important representations. Our framework is able to predict diseases and gives importance to different regions in an end-to-end manner, without using any surrogate models or ad-hoc mechanism. The model attributes the input features and makes decisions by the system without perturbation, which is easily implemented for clinical applications.

### 5.2. Investigation of findings

In our interpretability analysis, we detected disease-specific brain functional network disturbances mainly occurring in the left visual network, the left salience/ventral attention network

and the bilateral default mode network for ASD, whereas the right default mode network was the most vulnerable region for ADHD. Consistently, brain connectivity disruption in the visual and salience/ventral attention networks has also been reported in previous ASD studies (Lombardo et al., 2019; Marshall et al., 2020). Functional hypoconnectivity in the visual network has been found to be a key biomarker of ASD and may be critical for social communication development. Abnormalities in the attentional networks are known to be related to poor executive control (Keehn, Müller, & Townsend, 2013; Tang et al., 2020). The MDCN illustrated that the default mode network, comprising the prefrontal cortex, is associated with task-irrelevant mental processes, and is another remarkable brain regions that exhibits network reorganization in both ASD and ADHD (Harikumar, Evans, Dougherty, Carpenter, & Michael, 2021). Previous studies have indicated that the dorsal and medial prefrontal cortex may contribute to the development of the tendency to initiate joint attention, which is a critical milestone in early development and social learning (Bakeman & Adamson, 1984; Baldwin, 1995; Mundy, 2003). Social processing studies have found evidences of hypoactivation in the nodes of the "social brain", including the prefrontal cortex (Bush, 2011; Cubillo, Halari, Giampietro, Taylor, & Rubia, 2011; Dichter, 2022; Paloyelis, Mehta, Kuntsi, & Asherson, 2007). Since the prefrontal cortex has been implicated in social tasks (Schulte-Rüther et al., 2011), a disturbance in these networks may contribute to the atypical development of joint attention, social cognition and behavioral abnormalities (Cheng, Liu, Shi, & Yan, 2017; Mundy, 2003). Taken together, our findings couple with previous studies highlight the necessity of studying hypoconnectivity within the default mode network for both ASD and ADHD. (Anderson et al., 2014; Gilbert, Bird, Brindley, Frith, & Burgess, 2008; Hale et al., 2014; Minshew & Keller, 2010; Philip et al., 2012).

Interestingly, the dorsal and medial prefrontal cortex has been repeatedly detected in both ASD and ADHD (see Table 3: the right PFCdPFCm_3 with the adjusted p-value = 0.033 in ABIDE; the right PFCdPFCm_2 with the adjusted p-value = 0.027 in ADHD-200). This finding may be attributed to the similar symptomatology in children with ASD and ADHD, indicating that the two conditions might fall on the same continuum and are likely to share many elements of possible factors (Grzadzinski et al., 2011; Kern, Geier, Sykes, Geier, & Deth, 2015). A previous study (Yerys et al., 2019) found that reduced functional connectivity in the fronto-parietal and attention networks is linked to ADHD symptoms in ASD patients. Nevertheless, others have stated that functional abnormalities in the fronto striatal system (including the prefrontal cortex) are involved in both ASD and ADHD (Rommelse, Geurts, Franke, Buitelaar, & Hartman, 2011). Both functional and structural studies have shown abnormalities in this system in ASD and ADHD, as the system is responsible for adaptive responses (Durston, de Zeeuw, & Staal, 2009; Takarae, Minshew, Luna, & Sweeney, 2007). Our finding of consistent changes in the prefrontal cortex coincides with these studies and provides novel evidence for examining and comparing the neural signatures of both conditions.

### 5.3. Limitations and future work

This study has some limitations. The proposed interpretation module p-GradCAM, depends on the classification score returned from the neural network, which is not well suited to regression tasks. Nevertheless, extension to regression tasks (e.g., examining brain connectome-IQ relationship) is feasible if the regression values are obtained by transforming the classification logits with densely distributed regression values. Second, in this study, although we only evaluated the functional connectome, the proposed MDCN has practical application prospects for other connectome studies. Our previous studies have shown the potential use of MDMR in studying both functional and structural connectomes (Yang et al., 2021; Ye et al., 2019) in Alzheimer's disease and Parkinson's disease, while this can also be done by MDCN. In future works, we plan to evaluate the framework for discovering more types of connectivity-disease relationships and demonstrate its extensibility. Finally, compared to that of other deep learning methods, such as BrainNetCNN and BrainGNN, our proposed method is somewhat computationally expensive, where a two-stage training strategy is applied, and the graph training is repeated for each parcellation. However, we suggest that this is meaningful and acceptable because the current deep learning models for connectome studies are remain small. The trends of the increasing number of aggregated data and increasing model parameters are triggered in multiple areas, which might provide insights into brain connectome studies.

## 6. Conclusion

Exploring the complex associations between the brain connectome and clinical phenotypes is a challenging task in neuroscience. In this study, we propose a unified connectome-based deep learning framework and demonstrate its feasibility and generalization ability using two large datasets. Our approach not only outperforms other state-of-the-art methods of disease identification but also provides interpretable feature mapping individually for clinical guidance. The network reorganization detected by MDCN, primarily located in the dorsal and medial prefrontal cortex, is shared by both ASD and ADHD. Owing to the advantages of MDCN, we believe it is important to systematically integrate connectome mapping and graph learning to achieve better interpretable and accurate disease identification, and it is desirable to extend this framework in that direction in the future.

## CRediT authorship contribution statement

**Yanwu Yang:** Conceptualization, Methodology, Statistical analysis, Visualization, Writing – original draft & editing. **Chenfei Ye:** Conceptualization, Writing – review & editing, Supervision. **Ting Ma:** Research design, Writing – review & editing, Funding acquisition, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data are publicly available online at https://github.com/podismine/MDCN.

## Acknowledgments

This study is by grants from the National Natural Science Foundation of China (62276081 and 62106113), The National Key Research and Development Program of China (2021YFC2501202), the Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (NO: ZYYCXTD-C-202004), Basic Research Foundation of Shenzhen Science and Technology Stable Support Program (GXWD20201230155427003-20200822115709001).

## References


Anderson, A., Douglas, P. K., Kerr, W. T., Haynes, V. S., Yuille, A. L., Xie …, J., & Cohen, M. S. (2014). Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD. *Neuroimage*, *102*, 207–219.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 1278–1289.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. *Joint Attention: Its Origins and Role in Development*, *131*, 158.

Bargmann, C. I., & Marder, E. (2013). From the connectome to brain function. *Nature Methods*, *10*(6), 483–490.

Bellec, P., Benhajali, Y., Carbonell, F., Dansereau, C., Albouy, G., Pelland ., M., & Stip, E. (2015). Impact of the resolution of brain parcels on connectome-wide association studies in fMRI. *Neuroimage*, *123*, 212–228.

Bullmore, E. T., & Bassett, D. S. (2011). Brain graphs: graphical models of the human brain connectome. *Annual Review of Clinical Psychology*, *7*, 113–140.

Bush, G. (2011). Cingulate, frontal, and parietal cortical dysfunction in attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *69*(12), 1160–1167.

Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision*.

Cheng, J., Liu, A., Shi, M. Y., & Yan, Z. (2017). Disrupted glutamatergic transmission in prefrontal cortex contributes to behavioral abnormality in an animal model of ADHD. *Neuropsychopharmacology*, *42*(10), 2096–2104.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173.

Cubillo, A., Halari, R., Giampietro, V., Taylor, E., & Rubia, K. (2011). Fronto-striatal underactivation during interference inhibition and attention allocation in grown up children with attention deficit/hyperactivity disorder and persistent symptoms. *Psychiatry Research: Neuroimaging*, *193*(1), 17–27.

Dichter, G. S. (2022). Functional magnetic resonance imaging of autism spectrum disorders. *Dialogues in Clinical Neuroscience*.

Durston, S., de Zeeuw, P., & Staal, W. G. (2009). Imaging genetics in ADHD: a focus on cognitive control. *Neuroscience & Biobehavioral Reviews*, *33*(5), 674–689.

Feng, Y., You, H., Zhang, Z., Ji, R., & Gao, Y. (2019). Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nature Reviews Neuroscience*, *16*(3), 159–172.

Gilbert, S. J., Bird, G., Brindley, R., Frith, C. D., & Burgess, P. W. (2008). Atypical recruitment of medial prefrontal cortex in autism spectrum disorders: An fMRI study of two executive function tasks. *Neuropsychologia*, *46*(9), 2281–2291.

Grzadzinski, R., Di Martino, A., Brady, E., Mairena, M. A., O'Neale, M., Petkova ., E., & Castellanos, F. X. (2011). Examining autistic traits in children with ADHD: does the autism spectrum extend to adhd? *Journal of Autism and Developmental Disorders*, *41*(9), 1178–1191.

Hale, T. S., Kane, A. M., Kaminsky, O., Tung, K. L., Wiley, J. F., McGough ., J. J., & Kaplan, J. T. (2014). Visual network asymmetry and default mode network function in ADHD: an fMRI study. *Frontiers in Psychiatry*, *5*, 81.

Harikumar, A., Evans, D. W., Dougherty, C. C., Carpenter, K. L., & Michael, A. M. (2021). A review of the default mode network in autism spectrum disorders and attention deficit hyperactivity disorder. *Brain Connectivity*, *11*(4), 253–263.

Huang, F., Tan, E.-L., Yang, P., Huang, S., Ou-Yang, L., Cao ., J., & Lei, B. (2020). Self-weighted adaptive structure learning for ASD diagnosis via multi-template multi-center representation. *Medical Image Analysis*, *63*, Article 101662.

Jiang, H., Cao, P., Xu, M., Yang, J., & Zaiane, O. (2020). Hi-GCN: a hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. *Computers in Biology and Medicine*, *127*, Article 104096.

Jiang, J., Wei, Y., Feng, Y., Cao, J., & Gao, Y. (2019). Dynamic hypergraph neural networks. *IJCAI*.

Katarya, R., Sharma, P., Soni, N., & Rath, P. (2022). A review of interpretable deep learning for neurological disease classification. In *2022 8th international conference on advanced computing and communication systems*.

Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau ., R. E., & Hamarneh, G. (2017). BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *Neuroimage*, *146*, 1038–1049.

Keehn, B., Müller, R.-A., & Townsend, J. (2013). Atypical attentional networks and the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *37*(2), 164–183.

Kern, J. K., Geier, D. A., Sykes, L. K., Geier, M. R., & Deth, R. C. (2015). Are ASD and ADHD a continuum? A comparison of pathophysiological similarities between the disorders. *Journal of Attention Disorders*, *19*(9), 805–827.

Kong, Y., Gao, S., Yue, Y., Hou, Z., Shu, H., Xie ., C., & Yuan, Y. (2021). Spatio-temporal graph convolutional network for diagnosis and treatment response prediction of major depressive disorder from functional connectivity. *Human Brain Mapping*, *42*(12), 3922–3933.

Ktena, S. I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., & Rueckert, D. (2018). Metric learning with spectral graph convolutions on brain connectivity networks. *Neuroimage*, *169*, 431–442.

Lee, J. K., Amaral, D. G., Solomon, M., Rogers, S. J., Ozonoff, S., & Nordahl, C. W. (2020). Sex differences in the amygdala resting-state connectome of children with autism spectrum disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(3), 320–329.

Li, X., Dvornek, N. C., Zhou, Y., Zhuang, J., Ventola, P., & Duncan, J. S. (2019). Graph neural network for interpreting task-fmri biomarkers. In *International conference on medical image computing and computer-assisted intervention*.

Li, L., Jiang, H., Wen, G., Cao, P., Xu, M., Liu ., X., & Zaiane, O. (2021). TE-HI-GCN: an ensemble of transfer hierarchical graph convolutional networks for disorder diagnosis. *Neuroinformatics*, 1–23.

Li, L., Jiang, H., Wen, G., Cao, P., Xu, M., Liu ., X., & Zaiane, O. (2022). TE-HI-GCN: an ensemble of transfer hierarchical graph convolutional networks for disorder diagnosis. *Neuroinformatics*, *20*(2), 353–375.

Li, H., Satterthwaite, T. D., & Fan, Y. (2017). Large-scale sparse functional networks from resting state fMRI. *Neuroimage*, *156*, 1–13.

Li, W., van Tol, M. J., Li, M., Miao, W., Jiao, Y., Heinze ., H. J., & Walter, M. (2014). Regional specificity of sex effects on subcortical volumes across the lifespan in healthy aging. *Human Brain Mapping*, *35*(1), 238–247.

Lin, H.-Y., Ni, H.-C., Tseng, W.-Y. I., & Gau, S. S.-F. (2020). Characterizing intrinsic functional connectivity in relation to impaired self-regulation in intellectually able male youth with autism spectrum disorder. *Autism*, *24*(5), 1201–1216.

Liu, N., Feng, Q., & Hu, X. (2022). Interpretability in graph neural networks. In *Graph neural networks: foundations, frontiers, and applications* (pp. 121–147). Springer.

Lombardo, M. V., Eyler, L., Moore, A., Datko, M., Barnes, C. C., Cha ., D., & Pierce, K. (2019). Default mode-visual network hypoconnectivity in an autism subtype with pronounced social visual engagement difficulties. *Elife*, *8*, Article e47427.

Marshall, E., Nomi, J. S., Dirks, B., Romero, C., Kupis, L., Chang, C., & Uddin, L. Q. (2020). Coactivation pattern analysis reveals altered salience network dynamics in children with autism spectrum disorder. *Network Neuroscience*, *4*(4), 1219–1234.

Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., & Khudanpur, S. (2010). *Recurrent neural network based language model*. Interspeech.

Milham, M. P. (2012). Open neuroscience solutions for the connectome-wide association era. *Neuron*, *73*(2), 214–218.

Minshew, N. J., & Keller, T. A. (2010). The nature of brain dysfunction in autism: functional brain imaging studies. *Current Opinion in Neurology*, *23*(2), 124.

Mundy, P. (2003). Annotation: The neural basis of social impairments in autism: The role of the dorsal medial-frontal cortex and anterior cingulate system. *Journal of Child Psychology and Psychiatry*, *44*(6), 793–809.

Nt, H., & Maehara, T. (2019). Revisiting graph neural networks: All we have is low-pass filters. arXiv preprint arXiv:1905.09550.

Paloyelis, Y., Mehta, M. A., Kuntsi, J., & Asherson, P. (2007). Functional MRI in ADHD: a systematic literature review. *Expert Review of Neurotherapeutics*, *7*(10), 1337–1356.

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., & Rueckert, D. (2018a). Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease. *Medical Image Analysis*, *48*, 117–130.

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., & Rueckert, D. J. M. i. a. (2018b). Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease. *48*, 117–130.

Philip, R. C., Dauvermann, M. R., Whalley, H. C., Baynham, K., Lawrie, S. M., & Stanfield, A. C. (2012). A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders. *Neuroscience & Biobehavioral Reviews*, *36*(2), 901–942.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should i trust you? Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Rommelse, N. N., Geurts, H. M., Franke, B., Buitelaar, J. K., & Hartman, C. A. (2011). A review on cognitive and brain endophenotypes that may be common in autism spectrum disorder and attention-deficit/hyperactivity disorder and facilitate the search for pleiotropic genes. *Neuroscience & Biobehavioral Reviews*, *35*(6), 1363–1396.

Sak, H., Senior, A. W., & Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes ., A. J., & Yeo, B. T. (2018). Local–global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114.

Schulte-Rüther, M., Greimel, E., Markowitsch, H. J., Kamp-Becker, I., Remschmidt, H., Fink, G. R., & Piefke, M. (2011). Dysfunctions in brain networks supporting empathy: an fMRI study in adults with autism spectrum disorders. *Social Neuroscience*, *6*(1), 1–21.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*.

Sharma, A., Wolf, D. H., Ciric, R., Kable, J. W., Moore, T. M., Vandekar ., S. N., & Davatzikos, C. (2017). Common dimensional reward deficits across mood and psychotic disorders: a connectome-wide association study. *American Journal of Psychiatry*, *174*(7), 657–666.

Shehzad, Z., Kelly, C., Reiss, P. T., Craddock, R. C., Emerson, J. W., McMahon ., K., & Milham, M. P. (2014). A multivariate distance-based analytic framework for connectome-wide association studies. *Neuroimage*, *93*, 74–94.

Sheng, W., Cui, Q., Jiang, K., Chen, Y., Tang, Q., Wang ., C., & He, Z. (2022). Individual variation in brain network topology is linked to course of illness in major depressive disorder. *Cerebral Cortex*.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg ., H., & Flitney, D. E. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23*, S208–S219.

Song, X., Zhou, F., Frangi, A. F., Cao, J., Xiao, X., Lei ., Y., & Lei, B. (2021). Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction. *Medical Image Analysis*, *69*, Article 101947.

Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Computational Biology*, *1*(4), Article e42.

Takarae, Y., Minshew, N. J., Luna, B., & Sweeney, J. A. (2007). Atypical involvement of frontostriatal systems during sensorimotor control in autism. *Psychiatry Research: Neuroimaging*, *156*(2), 117–127.

Tang, S., Sun, N., Floris, D. L., Zhang, X., Di Martino, A., & Yeo, B. T. (2020). Reconciling dimensional and categorical models of autism heterogeneity: a brain connectomics and behavioral study. *Biological Psychiatry*, *87*(12), 1071–1082.

van den Heuvel, M. P., Scholtens, L. H., & Kahn, R. S. (2019). Multiscale neuroscience of psychiatric disorders. *Biological Psychiatry*, *86*(7), 512–522.

van den Heuvel, M. P., & Sporns, O. (2019). A cross-disorder connectome landscape of brain dysconnectivity. *Nature Reviews Neuroscience*, *20*(7), 435–446.

Yan, W., Qu, G., Hu, W., Abrol, A., Cai, B., Qiao ., C., & Calhoun, V. D. (2022). Deep learning in neuroimaging: Promises and challenges. *IEEE Signal Processing Magazine*, *39*(2), 87–98.

Yang, Z., Kelly, C., Castellanos, F. X., Leon, T., Milham, M. P., & Adler, L. A. (2016). Neural correlates of symptom improvement following stimulant treatment in adults with attention-deficit/hyperactivity disorder. *Journal of Child and Adolescent Psychopharmacology*, *26*(6), 527–536.

Yang, Y., Ye, C., Sun, J., Liang, L., Lv, H., Gao ., L., & Wu, T. (2021). Alteration of brain structural connectivity in progression of Parkinson's disease: a connectome-wide network analysis. *NeuroImage: Clinical*, *31*, Article 102715.

Ye, C., Mori, S., Chan, P., & Ma, T. (2019). Connectome-wide network analysis of white matter connectivity in Alzheimer's disease. *NeuroImage: Clinical*, *22*, Article 101690.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead ., M., & Polimeni, J. R. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*.

Yerys, B. E., Tunç, B., Satterthwaite, T. D., Antezana, L., Mosner, M. G., Bertollo ., J. R., & Herrington, J. D. (2019). Functional connectivity of frontoparietal and salience/ventral attention networks have independent associations with co-occurring attention-deficit/hyperactivity disorder symptoms in children with autism. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(4), 343–351.

Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Network-based statistic: identifying differences in brain networks. *Neuroimage*, *53*(4), 1197–1207.

Zhang, Y., Tetrel, L., Thirion, B., & Bellec, P. (2021). Functional annotation of human cognitive states using deep graph convolution. *Neuroimage*, *231*, Article 117847.

Zhao, K., Duka, B., Xie, H., Oathes, D. J., Calhoun, V., & Zhang, Y. (2022). A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD. *Neuroimage*, *246*, Article 118774.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.