Neuroinformatics
https://doi.org/10.1007/s12021-021-09548-1

ORIGINAL ARTICLE

TE-HI-GCN: An Ensemble of Transfer Hierarchical Graph Convolutional Networks for Disorder Diagnosis

Lanting Li[1,2] · Hao Jiang[1] · Guangqi Wen[1,2] · Peng Cao[1,2] · Mingyi Xu[1] · Xiaoli Liu[3] · Jinzhu Yang[1,2] · Osmar Zaiane[4]

Accepted: 14 August 2021

Abstract

Accurate diagnosis of psychiatric disorders plays a critical role in improving the quality of life for patients and potentially supports the development of new treatments. Graph convolutional networks (GCNs) are shown to be successful in modeling applications with graph structures. However, training an accurate GCNs model for brain networks faces several challenges, including high dimensional and noisy correlation in the brain networks, limited labeled training data, and depth limitation of GCN learning. Generalization and interpretability are important in developing predictive models for clinical diagnosis. To address these challenges, we proposed an ensemble framework involving hierarchical GCN and transfer learning for sparse brain networks, which allows GCN to capture the intrinsic correlation among the subjects and domains, to improve the network embedding learning for disease diagnosis. Extensive experiments on two real medical clinical applications: diagnosis of Autism spectrum disorder (ASD) and diagnosis of Alzheimer's disease (AD) on both the ADNI and ABIDE databases, showing the effectiveness of the proposed framework. We achieved state-of-the-art accuracy and AUC for AD/MCI and ASD/NC (Normal control) classification in comparison with studies that used functional connectivity as features or GCN models. The proposed TE-HI-GCN model achieves the best classification performance, leading to about 27.93% (31.38%) improvement for ASD and 16.86% (44.50%) for AD in terms of accuracy and AUC compared with the traditional GCN model. Moreover, the obtained clustering results show high correspondence with the previous neuroimaging derived evidence of within and between-networks biomarkers for ASD. The discovered subnetworks are used as evidence for the proposed TE-HI-GCN model. Furthermore, this work is the first attempt of transfer learning on the two related disorder domains to uncover the correlation among the two diseases with a transfer learning scheme.

Keywords Graph convolutional networks · Disorder disease diagnosis · Brain network · Resting-state fMRI · Transfer learning

Introduction

Autism spectrum disorder (ASD) (Bajestani et al., 2019; Heinsfeld et al., 2018) and Alzheimer's disease (AD) (Wang et al., 2018; Duc et al., 2020) are neurodevelopmental and neurodegenerative disorders respectively, with devastating effects not only on the individual but also the society. Neuroimaging has provided relevant information on the diagnostic status and disease progression of brain disorders. Resting-state fMRI images provide us blood-oxygenation-level-dependent (BOLD) signals as a neurophysiological index to probe brain activity, which has been applied to the diagnosis of ASD and AD. The rs-fMRI data has a complex structure, which is inherently represented as a network with a set of nodes and edges (Khosla et al., 2019; Wang et al., 2020). There has been evidence of network-level changes in the ASD brain compared to an NC (Normal control) brain. Therefore, many works focus on modeling the whole brain rs-fMRI as a network (Qi et al., 2015; Zhu et al., 2019) and extracting representation from the network (Khazaee et al., 2016; Mier & Mier, 2015). The commonly used features are calculated based on graph-theoretic analysis, such

✉ Peng Cao
caopeng@cse.neu.edu.cn

[1] Computer Science and Engineering, Northeastern University, Shenyang, China

[2] Key Laboratory of Intelligent Computing in Medical Image of Ministry of Education, Northeastern University, Shenyang, China

[3] Department of Chemical and Biomolecular Engineering, Faculty of Engineering, National University of Singapore, Singapore, Singapore

[4] Amii, University of Alberta, Edmonton, Alberta, Canada

Published online: 11 November 2021

as clustering coefficients and local clustering coefficients (Wang et al., 2010) based on the local connectivity patterns among brain regions. However, hand-crafted network features may not be precise enough to represent or characterize the brain networks (Chen et al., 2017; Guo et al., 2017).

In order to analyze the network data, a surge of network embedding (a.k.a. graph embedding or graph representation learning) methods have been proposed (Yue et al., 2020; Grover & Leskovec, 2016; Tang et al., 2015), where their goal is to automatically learn a low-dimensional feature representation for a network while maximally preserving the network structure. In recent years, the high-level feature representation of deep convolutional neural networks has been proven superior to hand-crafted low-level and mid-level features (Ebrahimighahnavieh et al., 2020; Lundervold & Lundervold, 2019; Litjens et al., 2017). However, convolutional neural networks and recurrent neural networks have mainly focused on the grid-structured inputs rather than network structure data. Kipf and Welling (2016) proposed graph convolutional networks (GCN) as an effective graph embedding model that naturally combines structure information and node features in the learning process. It has emerged to learn deep representations of graph-structured data and has shown to outperform other traditional relational learning methods (Zhou et al., 2020; Wu et al., 2020). Recent work has applied GCN on the functional network derived from rs-fMRI data to extract latent features from a graph (Parisot et al., 2017; Li et al., 2019; Ktena et al., 2018; Parisot et al., 2018). Inspired by these works, we focus particularly on GCN methods for analyzing the neuroimages of brain disorders prediction in an end-to-end fashion. However, at the current stage, the fMRI image classification via GCN models faces many challenges as follow:

Challenge 1: Noisy correlations in the brain network.

In the brain network, considering all the correlations may lead to the inclusion of noisy and spurious connections. The presence of noise in brain images is owing to the fact that measurement errors are likely to arise due to technological limitations, operator performance, equipment, environment, and other factors (Vaishali et al., 2015). Currently, Pearson's Correlation Coefficient (PCC) is the simplest and most widely-used method in constructing functional brain networks. However, the PCC tends to result in a brain network with dense connections. The both issues cause overfitting issues and increase computational complexity. Due to its high dimensionality and high noise levels, analysis of a large brain functional network may not be powerful enough and easy to interpret. Removing weak (potentially noisy) connections depends on a hardthreshold without enough flexibility. The prediction accuracy as well as reliable and explainable biomarkers still remain the key focus of brain networks research.

Challenge 2: Limited labeled training data.

Network embedding learning with GCN requires a large collection of training data. However, another challenge is that the amount of available labeled data is usually very small in the clinical application, which limits the classification performance.

Challenge 3: Depth limitation of GCN learning.

Li et al. (2018) and Li et al. (2019) recently studied the depth limitations of GCNs and showed that deep GCNs could cause over-smoothing, which results in features at vertices within each connected component converging to the same value. As a result, most state-of-the-art GCN models are not deeper than 3 or 4 layers. Oversmoothing has been assumed to be the major cause of a performance drop in GCNs.

All the issues hinder network embedding learning for GCN. The motivation for this work stems from the problem of training GCN on the brain networks data. In this work, we aim to construct a cleaned brain network and network embedding model in a jointly learning manner and improve the network embedding learning by exploiting the potential associations among the subjects and the clinical disease domains. Our main contributions to the brain network classification are summarized as follows:

1. Removing noisy correlations in the brain network.

The feature reduction method can eliminate the noisy features and thus improve computational efficiency, classification and interpretation of the results. Brain networks are organized across multiple spatial scales and also can be analyzed at topological (network) scales ranging from individual nodes to the network as a whole. Therefore, the multi-scale scheme can sparsify the brain networks by removing weak connections. At the same time, we incorporate a multi-graph clustering (MGC) into a GCN model to enhance the important connections and further remove the irrelevant connections with a supervision scheme. The multi-scale brain networks construction combined with clustering could generate more robust and biologically meaningful functional connectivity networks.

2. Exploiting the association within the subjects for GCN.

To exploit the association in the subjects, all the training samples are treated as networks, and the aim is to learn a network embedding for subjects by preserving both the topology structure within individual brain functional networks and the association among the global population network. The previous work on network embedding learning of brain functional networks considers each instance independently in the learning process, ignoring the association among instances. Incorporating and preserving the intrinsic data association can promote learning a better embedding of the brain functional network by capturing global information. Modeling the networks in a hierarchical fashion also increases the receptive field of graph convolutions and allows for training a deeper GCN.

3. Transfer learning across the relevant disorder domains.

The transfer learning technique is another good choice for dealing with limited data. However, it has received less attention in the disorder domain. To leverage the association in the clinical disease domain, we design a transfer learning framework for network data across neurological disorders. In medical applications, transfer learning is commonly performed by taking a standard ImageNet architecture along with its pre-trained weights and then fine-tuning it on the target task. However, the classification task in ImageNet and disorder diagnosis have considerable differences (Raghu et al., 2019). Given the marked differences between natural images and medical images, we hypothesize that transfer learning can achieve more powerful target models if the source models are built directly on the brain networks from relevant diseases (Zhou et al., 2019). Some previous works have demonstrated the aspects of similarity in etiology and pathology between autism and Alzheimer's disease (Nasrat et al., 2017; Khan et al., 2016; Eid & Eid, 2019). In our work, we conduct transfer learning with GCN for network data acquired from patients with different diseases (ASD and AD) and investigate the correlation in the related brain disorder domains in a computational framework. Such finding is crucial evidence for the generalization of existing knowledge across populations for early diagnosis and prognosis of brain disease diagnosis.

More specifically, we propose an **E**nsemble of **T**ransfer **HI**erarchical **G**raph **C**onvolutional **N**etworks, called TE-HI-GCN, to improve the performance of disease diagnosis with a limited amount of labeled data. The connectivity in brain networks is characterized at different levels to better study the multi-scale of brain networks. Selecting an appropriate network architecture for analyzing rs-fMRI data is not trivial. To achieve it, we first employ multiple thresholds to generate sparse connectivity networks to reflect different levels of the topological structure of the original connectivity networks. For each sparse network, we propose a hierarchical GCN (HI-GCN) framework for modeling the brain connectivity network and population network simultaneously to learn a network feature embedding while considering the network topology information and subject's association. Through the joint learning of HI-GCN, a high-level embedding of brain network representation can be effectively learned in an end-to-end fashion with global supervision such that the embedding learned is useful for classification. On the other hand, we propose a transfer learning scheme enabling HI-GCN to learn generic graph structural features by leveraging the commonality in two related domains. To transfer the appropriate knowledge for the network data avoiding negative transferring, the transfer learning is also carefully conducted on the multiple levels of topological structure in the original connectivity network. Finally, for final clinical decision-making, we construct an ensemble classifier from multiple HI-GCN as target-level representations, each of which is obtained by training and transferring on the multiple levels sparse connectivity network. Moreover, the multiple HI-GCN models trained on the networks with different sparsity levels can reduce the chance that negative transfer happens. Extensive experiments on two real medical clinical applications: diagnosis of ASD and diagnosis of AD, which demonstrates network embedding learning from exploring the data correlations and transferring from the related domains can improve prediction performance. The code is available at: https://github.com/llt1836/TE-HI-GCN.

Furthermore, the proposed method has the following desirable properties:

1. **General**: TE-HI-GCN is a general learning model, which may be useful in other medical or biochemical applications with network data.
2. **State-of-the-Art**: TE-HI-GCN outperforms previous methods on three established datasets.
3. **Interpretability**: TE-HI-GCN can identify the biomarkers of subnetworks, and to the best of our knowledge, this work is the first attempt of transfer learning on the two related disorder domains to uncover the correlation and knowledge transfer across brain disorder diseases.
4. **Robust**: TE-HI-GCN can handle the small data with high dimension noisy connections. Finally, we experimented with different atlases, proving evidence of the robustness of the proposed method.

The rest of the paper is organized as follows. In "Related Work", we present the related work. In "Preliminaries of Graph Convolutional Networks (GCN)", we provide an introduction of the fMRI network and GCN. A detailed mathematical formulation of TE-HI-GCN is provided in "An Ensemble of Transfer Hierarchical Graph Convolutional Networks for Disorder Diagnosis, TE-HI-GCN". In "Experiment", we conducted extensive experiments to verify the advantage of our method for the diagnosis of ASD and AD. The conclusion is drawn in "Conclusion".

Related Work

Researchers have started exploring the application of deep learning methods to the analysis of fMRI. A relatively recent trend is to exploit neural networks for graph-structured data, such as Graph Convolution Networks or BrainNetCNN (Kawahara et al., 2017), to make individual-level predictions on connectomes. Recently, there are some research works introducing GCNs into fMRI analytics. Ktena et al. (2018)

used the graph representations to model the functional connectivity derived from fMRI data between a set of ROIs and proposed to learn a graph similarity metric using a siamese graph convolutional neural network. The proposed framework operated in the graph spectral domain to evaluate the similarity between a pair of graphs. Their method demonstrated to perform tasks of classification between matching and non-matching graphs by evaluating the similarity metrics between different brain connectivity networks. Yao et al. (2019) proposed a multi-scale triplet graph convolutional network (MTGCN) for brain functional connectivity analysis with rs-fMRI data. They first employ multi-scale templates for coarse-to-fine ROI parcellation to construct multi-scale FCs for each subject. Then a triplet GCN model is developed to learn multi-scale graph representations of brain FC networks, followed by a weighted fusion scheme for classification. Li et al. (2019) proposed a generalizable GCN inductive learning model to more accurately classify ASD v.s. Normal Controls (NC). The proposed GCN integrates all the available connectivity, geometric, anatomic information and fMRI related parameters into graphs for deep learning, and the proposed classifier is based on graph isomorphism, which can be applied to multi-graphs with different nodes/edges (e.g. sub-graphs). Anirudh and Thiagarajan (2019) proposed a bootstrapped version of graph convolutional neural networks (G-CNNs) that utilize an ensemble of weakly trained G-CNNs and reduce the sensitivity of models on the choice of graph construction.

Deep learning methods for computer aided mental disorder diagnosis in the neuroimages may be lack of underlying neural anatomical or functional evidence. Another line of work is finding the biomarkers associated with disorder disease. It is helpful for understanding the underlying roots of the disorder and can lead to earlier diagnosis and more targeted treatment. Li and Duncan (2020) proposed BrainGNN, a graph neural network (GNN) framework to analyze functional magnetic resonance images (fMRI) and discover neurological biomarkers. BrainGNN involves ROI-selection pooling layers (R-pool) that highlight salient ROIs and topK pooling (TPK) loss combined with group-level consistency (GLC) loss as regularization terms to encourage reasonable ROI-selection and preserve either individual- or group-level patterns. Arslan et al. (2018) applied a GCN for the classification of sexes based on the brain functional connectivity matrix derived from task fMRI data. He proposed an activation-based approach to identify salient graph nodes using spectral convolutional neural networks. The work in (Li & Duncan, 2020) and (Arslan et al., 2018) identify the most indicative ROIs (brain regions) after the node embedding learning with a series of graph convolution inspired by recent findings which suggest that some ROIs are more indicative of predicting neurological disorders than the others.

Preliminaries of Graph Convolutional Networks (GCN)

Supervised learning of brain networks. We define an undirected graph for each subject, $\boldsymbol{N}_i = \{\boldsymbol{\mathcal{R}}_i, \boldsymbol{\mathcal{A}}_i\}$, where $\boldsymbol{\mathcal{R}}_i = \{r_i^1, ..., r_i^M\}$ is the set of M nodes, and $\boldsymbol{\mathcal{A}}_i \in R^{M\times M}$ is the adjacency matrix describing the network's connectivity in the i-th subject, where M is the number of ROI. Here $M = 116$. The embedding of each vertex in $\boldsymbol{\mathcal{R}}$ is learned during the GCN training, therefore the initial value of $\boldsymbol{\mathcal{R}}_i$ is set to be one. Given a network $\boldsymbol{N}$, to identify whether a subject has a certain brain disorder can be regarded as a graph classification task. In the graph classification setting, we have a set of graphs $\{\boldsymbol{N}_1, \dots, \boldsymbol{N}_D\}$, where D is the size of dataset. Each graph $\boldsymbol{N}_i$ is associated with a label y_i.

Graph convolutional networks (GCN). Graph convolutional neural networks (GCN) aim to extend the data representation and classification capabilities of convolutional neural networks, which are highly effective for signals defined on regular Euclidean domains, e.g. image and audio signals, to irregular, graph-structured data defined on non-Euclidean domains. The graph convolution is employed directly on graph structured data to extract highly meaningful patterns and features in the space domain. Formally, given an adjacency matrix $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{M\times M}$, GCN is stacked by several convolutional layers can be written as:

$$\boldsymbol{E}^{(l+1)} = \mathbf{ReLu}(\tilde{\boldsymbol{D}}^{-1/2}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-1/2}\boldsymbol{E}^{(l)}\boldsymbol{W}^{(l)}), \tag{1}$$

where $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_n$, $\tilde{\boldsymbol{D}}_{ii} = \sum_j \tilde{\boldsymbol{A}}_{ij}$, $\boldsymbol{W}$ is a trainable weight matrix, $\boldsymbol{E}^{(l+1)}$ are the node embeddings computed after l steps of the GCN, and the node embeddings $\boldsymbol{E}^{(l)}$ generated from the previous message-passing step.

GCN can be considered as a Laplacian smoothing operator for node features over graph structures. The architecture of GCN consists of a series of convolutional layers, each followed by Rectified Linear Unit (ReLU) activation functions to increase non-linearity. The first hidden layer $\boldsymbol{E}^{(0)}$ is a set of the input original node features. All layers share the same adjacency matrix. A full GCN run L iterations of Eq. (1) to generate the final output node embeddings, $\boldsymbol{E}^{(L)}$.

An Ensemble of Transfer Hierarchical Graph Convolutional Networks for Disorder Diagnosis, TE-HI-GCN

This section starts with the architecture overview of our proposed predict-refine model, TE-HI-GCN in "Overview of Network Architecture of TE-HI-GCN". We describe the sparse brain network construction firstly in "Sparse Brain Networks Construction" followed by the details of our newly designed HI-GCN and transfer learning module in "Hierarchical GCN" and "Transfer Learning for GCN".

Overview of Network Architecture of TE-HI-GCN

First, we design an efficient multi-scale brain network learning framework in order to better understand the brain activity in the multi-level brain network (Betzel & Bassett, 2017). The proposed architecture is shown in Fig. 1. It consists of two components: the HI-GCN component for network embedding learning and the transfer learning component for knowledge transferring. Specifically, we first apply multiple thresholds to generate multiple thresholded connectivity networks to reflect different levels of topological structure of the original connectivity network. (Here, different thresholds determine their corresponding different levels of topological structure). In our study, the range of sparsity level in the brain networks is set to [0.05, 0.5].

Then, for each threshold value, we derive a sparse network of each subject and the corresponding population network. In order to better study the multi-scale of the brain network, the population network is also constructed by all the subjects with the same sparsity level. At each level, both hierarchical representation learning and transfer learning are trained and conducted on the sparse brain network data to guide the training of GCN. With the cooperation of two components, GCN can learn a discriminative network representation for the brain network, thus enabling improving network classification performance. However, it is difficult to make a principled choice of threshold values. Different thresholds determine their corresponding different levels of topological structure. Therefore, it is important to identify the optimal trade-off between the information gain by the removal of noisy edges and the loss due to the removal of potentially useful weak edges. Rather than optimize for the best threshold value, we adopt an ensemble classification strategy over a range of thresholds, a simple and effective fusion method with voting, to combine the results on multi-scale topological information in the brain network for clinical decision making. The ensemble method combines multiple models in order to get a better and more comprehensive generalized model. The diversity of the proposed TE-HI-GCN comes from the different training network data with different sparsity level representations. Therefore, our TE-HI-GCN produces different target-level representations of the brain networks, learned from possibly different sparse brain networks data. By exploiting multiple classifiers in an ensemble manner, we also expect the ensemble network can overcome the prediction noise of the predictive model and edge noise of the brain networks.

Sparse Brain Networks Construction

The construction of the brain network from fMRI involves two steps which are shown in Fig. 2. At first, the mean time series for a set of regions extracted from the automated anatomical labeling (AAL) atlas dividing the brain into 116 regions according to structural criteria (Tzourio-Mazoyer et al., 2002) are computed and normalised to zero mean and unit variance. Then, we compute the region-to-region brain correlations by Pearson's correlation coefficient (PCC).

$$Q(r_i, r_j) = \frac{Cov(v_i, v_j)}{\sigma_{v_i}, \sigma_{v_j}} \tag{2}$$

where $Cov(v_i, v_j)$ is the cross covariance between v_i and v_j, and σ_v denotes the standard deviation of v.

In a brain network, considering all the correlations may lead to the inclusion of weak and spurious connections. The

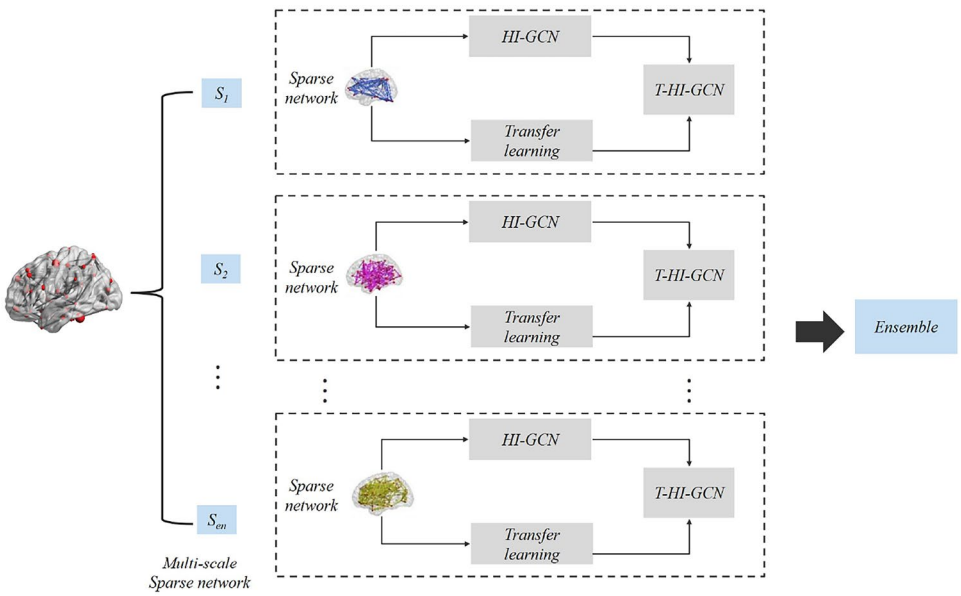

Fig. 1 The architecture of the proposed TE-HI-GCN model for brain network classification

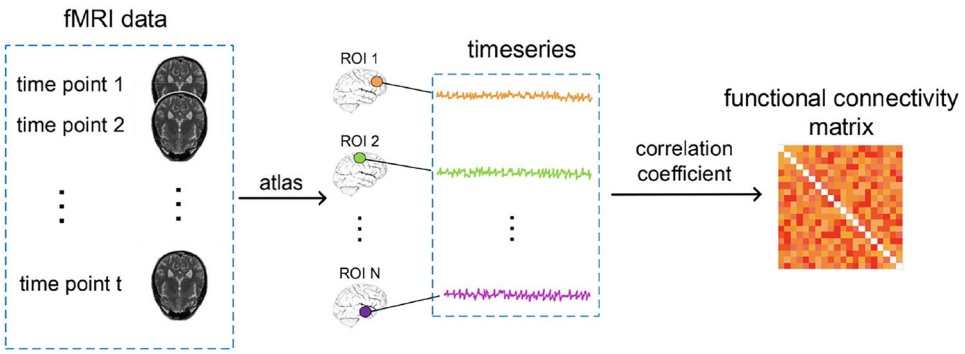

Fig. 2 The procedure of brain FC network construction

weak connections that are most influenced by experimental noise need to be removed for further analysis. The sparse representation-based brain network construction methods with thresholding strategies could generate more robust and biologically meaningful functional connectivity networks. We employ a thresholding operator to alleviate noises and outliers for constructing the brain networks and generating multiple thresholded connectivity networks, to reflect different levels of topological structure of the original connectivity network. Specifically, we use the threshold τ to filter noisy edges, and the operation can be written as

$$Q(r_i, r_j) = \begin{cases} Q(r_i, r_j) & \textbf{, if} \quad Q(r_i, r_j) \geq \tau \\ 0 & \textbf{, else} \end{cases} \tag{3}$$

Finally, we ensemble all the predictions to determine the predicted class using a normalized majority voting strategy.

Hierarchical GCN

In such a setting of the population analysis, each subject acquisition is represented by a node, and pairwise similarities are modeled via edges connecting the nodes. Given a collection of images modeled as graphs $\boldsymbol{N}_i$ and the associated label y_i, we construct a global population network $\hat{\boldsymbol{N}} = \left\{\hat{\boldsymbol{R}}, \hat{\boldsymbol{A}}\right\}$, where $\hat{\boldsymbol{A}}$ is the adjacency matrix describing the pairwise similarities between each pair of subjects with brain networks. Each subject is represented by a vertex $\hat{r}$ and is associated with a network data. This leads to a hierarchical graph in which a set of graph instances are interconnected via edges. This is a very expressive data representation, as it considers the relationship between graph instances rather than treating them independently. The definition of the graph's edges is critical in order to capture the underlying structure of the data and explain the similarities between each pair of the $\boldsymbol{N}$. We employ a graph kernel to estimate the $\hat{\boldsymbol{A}}(\boldsymbol{N}_i, \boldsymbol{N}_j)$ between two network inputs of subjects. The diagnosis with brain functional networks is a typical graph classification problem where brain networks are inputs and the predictions of the clinical stauts (i.e. patient with a disorder or normal control) are outputs. The aim is to learn the most essential embedding by taking full advantage of the correlation and structure within the graph and accurately predict the label of a given network.

The procedure of the embedding learning of the brain functional network is shown in Fig. 3. It includes two phrases:

1. f-GCN: learning the latent embedding representation of graph instance based on each ROI's connectivity into a meaningful low-dimensional representation for each brain network instance. The framework jointly optimizes the two parts: multi-graph clustering for correlation reduction and embedding learning with graph convolutions in a unified framework. The f-GCN model produces the embedding $\boldsymbol{E}$ for all network instances, then the learned embedding is fed to the second model (p-GCN).
2. p-GCN: further learning the graph embedding by message passing according to the network embedding $\boldsymbol{E}$ describing each subject and the adjacent matrix between samples $\hat{\boldsymbol{A}}$. As shown in Fig. 3, the input layer of p-GCN is defined as: $\hat{\boldsymbol{E}}^0 = \boldsymbol{E}^L$, where $\boldsymbol{E}^L$ is the set of node embedding features learned by f-GCN. The main idea is to generate a node $\hat{\boldsymbol{e}}$ representation by aggregating its own features $\boldsymbol{e}_i$ and neighbors' features $\boldsymbol{e}_j$, where $j \in \textbf{Neighbor}(i)$. p-GCN also stacks multiple graph convolutional layers to extract high-level node representations. The model inductively learns node representation by recursively aggregating and transforming feature vectors of its neighboring subjects. Finally, the p-GCN outputs a matrix $\hat{\boldsymbol{E}} \in R^{D \times \hat{P}}$, where the d-th row describes a latent representation of brain network from the d-th subject, and $\hat{P}$ is the conditionality of the final network embedding. Intuitively, and $\hat{\boldsymbol{E}}$ can be used as features for the tasks of brain disorder disease diagnosis.

The goal of HI-GCN for the graph classification task is to learn a nonlinear mapping from a brain network to an embedding vector. Note that both f-GCN and p-GCN are

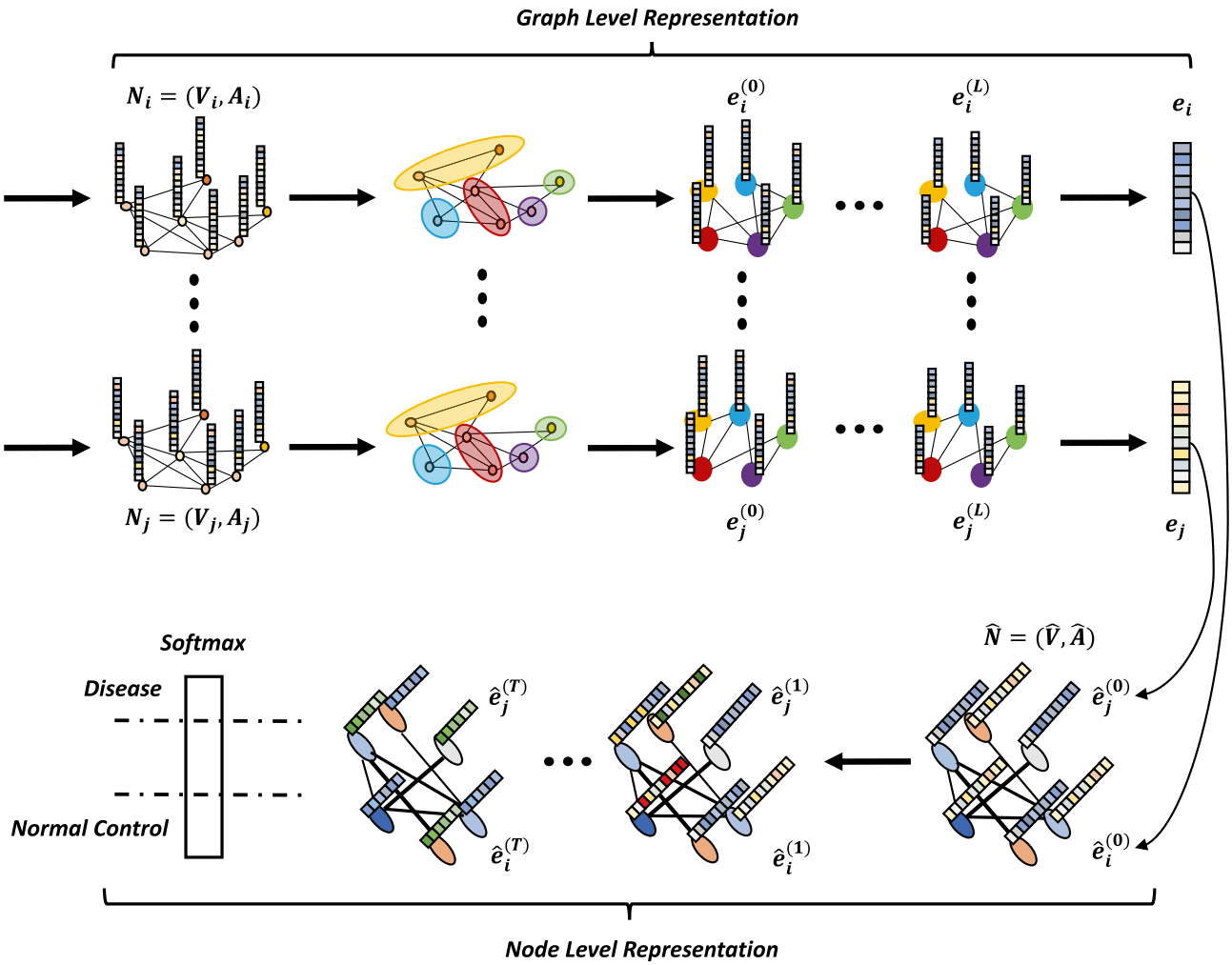

Fig. 3 An illustration of the procedure of network embedding learning in HI-GCN

jointly updated via backpropagation. The procedure is defined as:

$$\mathbf{HI-GCN}: N \rightarrow [\hat{e}, \hat{y}], \tag{4}$$

which involves two functions:

$$\mathbf{f-GCN}(N) = e; \qquad \mathbf{p-GCN}(e, \hat{\mathcal{A}}) = [\hat{e}, \hat{y}] \tag{5}$$

In the next subsection, we introduce details about the two parts of Hi-GCN respectively.

f-GCN: Learning the Brain Network Embedding with MGC

In a brain network, the dimension of functional connectivity could be relatively large and thus not very discriminant. The noisy connections that are most influenced by experimental noise need to be removed for further analysis. In this study, we develop a multi-graph clustering (Tang et al., 2009) based functional connectivity reduction strategy to obtain the clusters as supernodes and remove the noisy connections for diagnosing disorder diseases. The number of clusters is controlled as a hyperparameter. Our objective is well-motivated: by better reducing the noisy correlation edges with a multi-graph clustering, a better brain network can be learned.

Multi-graph clustering (MGC) aims to improve clustering accuracy by leveraging information from different domains. It has been shown to be extremely effective for achieving better clustering results than single graph-based clustering algorithms. One natural model for unsupervised graph clustering is to approximate the given graph through a low-rank matrix factorization $\boldsymbol{A} \approx \boldsymbol{F}^T \boldsymbol{A}^s \boldsymbol{F}$, where $\boldsymbol{\mathcal{F}}$ is an $M \times C$ matrix, and $\boldsymbol{\mathcal{A}}^s$ is an $C \times C$ symmetric matrix, C indicates the number of clusters. Given multiple graphs (brain networks), the underlying clustering $\boldsymbol{\mathcal{F}}$ is shared among graphs. The matrix $\boldsymbol{\mathcal{F}}$ to be optimized plays two roles: 1). Each item f_{ij} can be interpreted as the membership of the node i to the supernode S_j. 2). $\boldsymbol{\mathcal{F}}_i$ is assigned a value indicating how important

node i is for the prediction task. With MGC, we can obtain a set of clusters as supernodes $S_1, S_2, ..., S_C$ and the weighted adjacency matrix of the supergraph: $\boldsymbol{\mathcal{A}}^s = \boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{F}}^T$. With the supernodes $S_1, S_2, ..., S_C$ and the weighted adjacency matrix $\boldsymbol{\mathcal{A}}^s$, a coarsened graph is constructed. Then, three graph convolutional layers are stacked to learn the graph structure using the weighted adjacency matrix $\boldsymbol{\mathcal{A}}^s$.

Different from traditional clustering which groups the similar nodes, the aim of MGC is to hide the noisy connectivity by grouping them into a supernode, thus highlighting the indicative edges connecting the supernodes. In other words, the weight of functional connections connecting the nodes crossing different clusters is enhanced, whereas the nodes within clusters and their connections are removed.

Most learning approaches treat clustering and classification separately (i.e., sequentially), but recent research has shown that optimizing the two tasks jointly can substantially improve the performance of both. The question then is how to simultaneously learn the MGC and the GNN for time series in an end-to-end framework. We incorporate the grouping multi-graph clustering into the GCN model to reduce dimensionality as part of our end-to-end neural network model. Thus, the node clustering can be combined and blended with the graph convolution and classification with a supervision scheme.

p-GCN: Learning the Brain Network Embedding with Subject Correlation

The graph-based method has a wide range of applications. The data samples are the nodes of the graph, and the data relationship corresponds to the edges on the graph. The hypothesis of graph based classification is the importance of contextual pairwise information for the classification. As with many graph algorithms, the adjacency matrix encodes the pairwise relationship for training and test data. The learning of the model, as well as the embedding, is performed for both data simultaneously. These methods are transductive. For many applications, however, test data may not be readily available because the graph may constantly be expanding with new vertices. Such scenarios require an inductive scheme that learns a model from only a training set of vertices and generalizes well to any unseen instance. In our work, the graph-based learning with p-GCN is conducted in an inductive setting. In this paradigm, the training samples are represented as nodes in the population network $\hat{\boldsymbol{A}}$. The graph-based learning of p-GCN consists of two steps: graph construction and inference.

Graph construction: The definition of the graph's edges is critical in order to capture the underlying structure of the data and explain the similarities between the feature vectors. We employ a graph kernel to directly measure the topological similarity between functional connectivity networks. The graph kernel is one kind of kernel constructed on graphs that measures the topological similarity between graphs. More formally, given a pair of networks $\boldsymbol{N}_i$ and $\boldsymbol{N}_j$, a graph kernel can be defined as $\hat{A}(\boldsymbol{N}_i, \boldsymbol{N}_j) = \left\langle \phi(\boldsymbol{N}_i), \phi(\boldsymbol{N}_j) \right\rangle$, which takes into account the topology of networks $\boldsymbol{N}_i$ and $\boldsymbol{N}_j$.

In our work, we compute the similarity between the structure of two brain networks directly rather than the embeddings, and the similarity score between a pair of brain networks $\boldsymbol{N}_i$ and $\boldsymbol{N}_j$ is denoted by $\hat{A}(\boldsymbol{N}_i, \boldsymbol{N}_j)$. Kernel methods have the desirable property that they do not rely on explicitly characterizing the vector representation $\phi(x)$ of data points in the feature space induced by a kernel function but access data only via the Gram matrix $\mathcal{K}$. In this setting, a kernel $k : \mathbb{G} \times \mathbb{G} \to \mathbb{R}$ is called a graph kernel, which can capture the inherent similarity in the graph structure and is reasonably efficient to evaluate. Distances between instances with the q-th kernel function are calculated and are defined as $\mathcal{K}_q(r_a^i, r_b^i)$, where $r_a^i = \sum_u^M \hat{A}_i(a, u)$, indicating the local topology of nodes. We assume that the relation structures of brain networks belonging to the same class are relatively more similar, while those belonging to different classes are relatively more dissimilar. Like kernels on vector spaces, graph kernels can be calculated implicitly by computing $\mathcal{K}$. If the RBF kernel function is chosen, then the distance between instances is calculated as: $\mathcal{K}$, where σ is a kernel parameter.

To capture the similarity among networks, the similarity between networks $\boldsymbol{N}_i$ and $\boldsymbol{N}_j$ is calculated as:

$$S_I(\boldsymbol{N}_i, \boldsymbol{N}_j) = \frac{\sum_{a=1}^{M} \sum_{b=1}^{M} w_a^i w_b^j \mathcal{K}(\boldsymbol{x}_a^i, \boldsymbol{x}_b^j)}{\sum_M^{n_i} w_a^i \sum_{b=1}^{M} w_b^j} \tag{6}$$

where $w_a^i = \frac{1}{\sum_{u=1}^{M} \mathcal{K}(\boldsymbol{x}_a^i, \boldsymbol{x}_u^i)}$ is associated with each brain region r_a^i in $\boldsymbol{N}_i$ with the q-th kernel function.

Finally, a joint loss is trained for obtaining the clustering and classification as follow:

$$\min_{\boldsymbol{W}_f, \boldsymbol{W}_p, \boldsymbol{\mathcal{F}}} \boldsymbol{L} = \boldsymbol{L}_{CE}(\boldsymbol{W}_f, \boldsymbol{W}_p, \boldsymbol{\mathcal{F}}) + \lambda_1 \boldsymbol{L}_{otho}(\boldsymbol{\mathcal{F}}) + \lambda_2 \boldsymbol{L}_{bal}(\boldsymbol{\mathcal{F}}) + \lambda_3 \boldsymbol{L}_{pos}(\boldsymbol{\mathcal{F}}) \tag{7}$$

where $\boldsymbol{W}_f$ and $\boldsymbol{W}_p$ are the weight parameters of f-GCN and p-GCN, $\boldsymbol{\mathcal{F}}$ are the MGC matrix, λ_1, λ_2 and λ_3 are all positive parameters which control contributions of multiple regularization, respectively, $\boldsymbol{L}_{otho}$ is orthogonal regularization to penalize the off-diagonal elements of $\boldsymbol{\mathcal{F}}^T\boldsymbol{\mathcal{F}}$: $\boldsymbol{L}_{otho} = \left\| \boldsymbol{\mathcal{F}}^T\boldsymbol{\mathcal{F}} - diag(diag(\boldsymbol{\mathcal{F}}^T\boldsymbol{\mathcal{F}})) \right\|_F$, $\boldsymbol{L}_{bal}$ is a balancing regularization to achieve a balanced clustering: $\boldsymbol{L}_{bal} = Var(diag(\boldsymbol{\mathcal{F}}^T\boldsymbol{\mathcal{F}}))$, where $Var(\cdot)$ means variance, $\boldsymbol{L}_{pos}$ is to guarantee the value of $\boldsymbol{\mathcal{F}}$ is positive value.

Inference: When predicting previously unseen data $\boldsymbol{x}_i$, the network embedding $\boldsymbol{e}_i$ is generated by f-GCN at first. Then, a fixed-size set of neighbors of the unseen sample from the training set is obtained, the aim of which is to align the newly observed nodes to the training embeddings that

the algorithm has already optimized. The local adjacency matrix of the $\boldsymbol{x}_i$ and its training neighbors is indicated as $\hat{\mathcal{A}}_i$. Then, node embedding and predicated label are generated by applying the learned p-GCN model according to:

$$\hat{y}_i = \mathbf{p} - \mathbf{GCN}(\boldsymbol{e}_i, \hat{\boldsymbol{\mathcal{A}}}_i) \tag{8}$$

Transfer Learning for GCN

A sufficient number of training samples is necessary for training deep learning models. Unfortunately, larger datasets in medicine are often relatively small. However, this assumption can be relaxed if transfer learning is applied in conjunction with deep learning. To overcome this problem, transfer learning has recently been explored in various medical imaging applications. Through applying the transfer learning on the two datasets, we explore the hypothesis that the diagnosis model of brain disorders with GCN can be transferred across relevant diseases.

Define a network domain as $\boldsymbol{\Delta} = \left\{\hat{\boldsymbol{N}}, f(\hat{\boldsymbol{N}})\right\}$, which includes an population network $\hat{\boldsymbol{N}}$ constructed by all data and a function $f(\hat{\boldsymbol{N}})$ for the graph classification task. Then, the domain can be represented by $\boldsymbol{\Delta}_{ASD} = \left\{\hat{\boldsymbol{N}}_{ASD}, f_{ASD}(\hat{\boldsymbol{N}}_{ASD})\right\}$ and $\boldsymbol{\Delta}_{AD} = \left\{\hat{\boldsymbol{N}}_{AD}, f_{AD}(\hat{\boldsymbol{N}}_{AD})\right\}$, respectively. Transfer learning aims to boost the generalization capability of the predictive function through the transfer of knowledge from $\boldsymbol{\Delta}_{ASD}$ or $\boldsymbol{\Delta}_{AD}$ with its task f_{ASD} or f_{AD}. In transfer learning scenarios of this work, the domain and task are different. The task f_{AD} is to discriminate AD from MCI, whereas the task f_{ASD} is to classify subjects suffering from ASD from healthy control subjects.

We developed a novel strategy for the transfer GCN method to consider the different topological structure in both brain network level and population network level at the same time. Note that with pre-trained HI-GCN trained on the multiple source brain networks with different levels of topological structure, we fine-tune it on the target brain networks with corresponding sparsity level. Figure 4 shows the different pre-training schemes across two diseases: node clustering pretraining (MGC module in f-GCN) pretraining, node-level feature learning (GCN module in f-GCN) pretraining, the whole f-GCN pretraining, graph-level embedding learning (p-GCN) pretraining and the ensemble pretraining of E-HI-GCN with multi-sparsity level.

Implementation Details

The whole model is optimized in an end-to-end fully supervised manner. The hyperparameters in Eq. 7 are tuned empirically, which yields the best performance. The parameter setting of our model is shown in Table 1. For all the HI-GCN models, we choose full-batch training. The whole framework was built on PyTorch with GeForce RTX 3090 GPU for all our experiments.

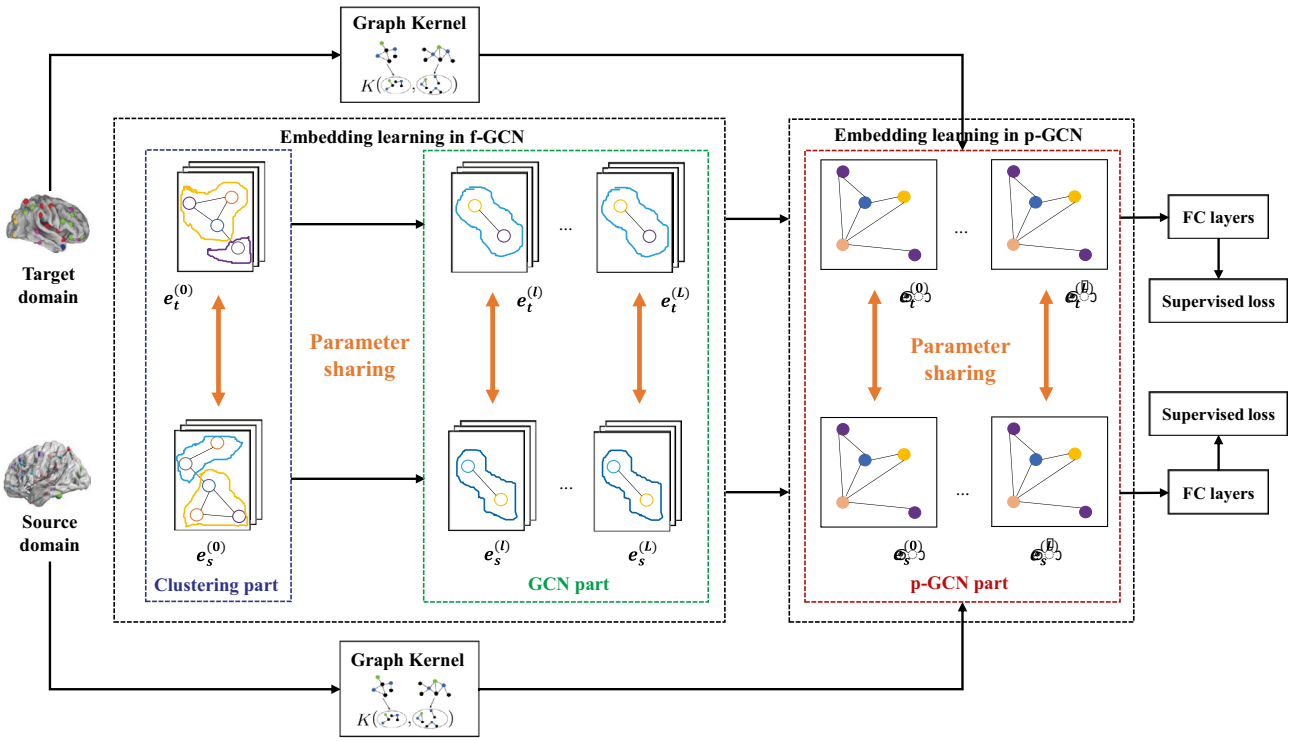

Fig. 4 The illustration of the transfer learning for HI-GCN

Table 1 The parameter settings of network training of TE-HI-GCN

parameter name	parameters
learning rate	0.001
learning rate schedule	CosineLR(Cosine Annealing with Warm Restarts)
T_max in CosineLR	50
batch size of f-GCN	32
Training iterations of f-GCN	300
Training iterations of E-HI-GCN for ABIDE & ADNI	2000 & 300
Number of clustering	5

The previous studies have found that positive and negative correlation connectivities have different contributions. Therefore, all the correlation connectivities are first split into positive and negative networks, each of which is performed clustering and graph convolution independently. At last, two embeddings are flattened and fed into the fully connected layer.

Experiment

In this section, we conducted several sets of comparative experiments and rigorously analyzed our results on both ABIDE and ADNI data sets. Next, we briefly introduce these comparative experiments for them. The experimental settings and results will be described in detail in the next subsections. In our experiments with the ADNI cohort, we validated the effectiveness of the proposed method by achieving the highest diagnostic accuracies in two classification tasks. We also rigorously analyzed our results and compared them with the previous studies on the ADNI and ABDIE cohort in the literature.

Databases and Preprocessing

We apply our model on two large and challenging databases for binary classification tasks. The ABIDE database (Autism Brain Imaging Data Exchange database investigates the neural basis of autism) (Di Martino et al., 2014) aggregates data from different 17 acquisition sites and openly shares rs-fMRI and phenotypic data of 1112 subjects. In this work, we used data from the ABIDE preprocessed connectome project (PCP) (Craddock et al., 2013) using the Configurable Pipeline for the Analysis of Connectomes (CPAC). The detail procession of PCP and CPAC can be referenced in (Friston et al., 1994; Fox et al., 2005; Lund et al., 2005; Behzadi et al., 2007). Bandpass filtering (0.01–10 Hz) and global signal correction was used in our analysis. After the preprocessing, we obtained 871 high quality fMRI time series phenotypic information, comprising 403 individuals with ASD and 468 normal controls.

The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. We focus on using rs-fMRI to discriminate individuals with Mild Cognitive Impairment (MCI) from individuals diagnosed with Alzheimer's Disease (AD). We select the same set of 133 subjects used in (Dadi et al., 2019), comprising individuals with 99 MCI and 34 diagnosed with Alzheimer's Disease (AD).

Evaluating the Effectiveness of our TE-HI-GCN

In this section, we conduct an empirical evaluation for the proposed methods by comparing the traditional method with network connectivity features and GCN as baseline methods on two publicly available rest-fMRI datasets: ABIDE and ADNI. For the ADNI dataset, the class of AD (minority) is considered as positive class and the class of MCI (majority) is considered as negative class.

Network connectivity Feature (NCF) (Abraham et al., 2017): the feature vector from brain networks is extracted by vectoring the functional connectivity matrix. Specifically, the upper triangle values in the functional connectivity matrix are extracted and flattened to a vector of features. In the connectivity matrix, there exist a large number of low level features (i.e., $\frac{M\times(M-1)}{2}$, where M is the total number of ROIs). Moreover, Recursive Feature Elimination (RFE) is used to feature selection. For a fair comparison, the amount of features selected is equal to the dimension of the embedding vector generated by f-GCN. At last, a ridge classifier is trained on the extracted feature vector.

GCN: we choose Eigenpooling GCN (Ma et al., 2019) as a baseline method, which is an end-to-end trainable graph pooling method producing hierarchical representations of graphs.

T-HI-GCN is a transfer learning method based on HI-GCN, which is a single classifier working on the fully connected brain networks.

E-HI-GCN is an ensemble of HI-GCN, each of which is trained on the different sparsity level brain networks.

In our systematic study, we find that pre-training does not always help. Both T-HI-GCN and TE-HI-GCN with the best transfer settings are shown in Tables 6 and 7. The ablation study to investigate the effect of each component pretraining are discussed in "The Effectiveness of Ensemble Learning"

We report the results of disease diagnosis of ASD and AD on the ABIDE and ADNI datasets in Tables 2 and 3. Our systematic study suggests the following trends:

Table 2 Performance comparison of various methods on ABIDE. The best results are bold

Methods	ACC	AUC	Precision	Sensitivity	F1
NCF	0.586±0.003	0.583±0.003	0.584±0.003	0.583±0.003	0.583±0.003
GCN	0.598±0.003	0.580±0.004	0.676±0.004	0.432±0.014	0.527±0.005
f-GCN	0.612±0.004	0.609±0.004	0.652±0.008	0.660±0.034	0.634±0.010
HI-GCN	0.672±0.004	0.666±0.004	0.682±0.005	0.725±0.005	0.710±0.003
T-HI-GCN	0.666±0.005	0.661±0.005	0.680 ±0.005	0.710 ±0.006	0.701±0.004
E-HI-GCN	0.735±0.001	0.750±0.001	0.745±0.003	0.720±0.007	0.745±0.001
TE-HI-GCN	**0.765±0.003**	**0.762±0.003**	**0.779±0.005**	**0.799±0.009**	**0.784±0.003**

Observation (1): From Tables 2 and 3, we can see that the proposed methods consistently achieve better classification performance than the competing methods on these two tasks in terms of accuracy (ACC) and AUC, which demonstrates the effectiveness of our TE-HI-GCN method. With the student's t-test at a level of 0.05, our proposed methods including TE-HI-GCN, E-HI-GCN,T-HI-GCN and HI-GCN significantly outperform GCN in most cases except the result on the ADNI dataset in terms of ACC. More specifically, it can be seen that the proposed TE-HI-GCN model achieves the best classification performance with an ACC (AUC) of 0.765 (0.762) for ASD and 0.894 (0.893) for AD. It has a clear impact on the quality of the predictions, leading to about 27.93%(31.38%) improvement for ABIDE and 16.85%(44.50%) for ADNI in terms of accuracy and AUC compared with the traditional GCN model. It means that our proposed TE-HI-GCN is able to explore the potential association in the samples and domains.

Observation (2): It can be observed that the performance of Eigenpooling GCN are poor, especially when the dataset size is small on ADNI. The results indicate that the traditional GCN cannot handle the brain networks classification problem well due to noisy correlations in the brain networks. Both f-GCN and HI-GCN consistently perform better than the Eigenpooling GCN. It further confirms that the noisy correlation reduction in f-GCN is effective for learning a clean network structure. Compared with f-GCN, HI-GCN performs the graph embedding learning from a hierarchical perspective considering the structure in individual brain network and the subject's correlation in the global population network, which can capture the essential embedding features to improve the classification performance of disease diagnosis.

Observation (3): It is worth noting that the proposed E-HI-GCN achieves the second best performance for ASD diagnosis in terms of ACC and AUC, respectively. Besides, it actually further confirms our finding that a single HI-GCN with a full connected network is worse than E-HI-GCN with the ensemble of sparse networks due to the inappropriate brain network construction. Incorporating ensemble learning allowed improving the classification performance further, as suggested by the results achieved by the E-HI-GCN and TE-HI-GCN models. The construction of multiple scale brain networks benefits the single GCN model by exploring different views of the brain network.

Observation (4): In this present study, we endeavored to examine whether our model trained on a specific population (AD/MCI or ASD/Normal control) is generalizable to other populations for the diagnosis of ASD/Normal control or AD/MCI. Experiments on the two datasets show that the pre-training strategy with appropriate setting achieves consistently better performance than the models from scratch. Based on the observations above, it can be concluded that the diagnosis model of brain disorders with GCN can be transferred across relevant diseases. It also confirmed our initial hypothesis about the association existing between ASD and AD on the brain networks. In addition, the strategy is able to partially overcome the overfitting problem caused by the limited amount of data in ADNI. The results obtained on the diagnosis task of AD in ADNI show a larger increase in performance with the proposed TE-HI-GCN over the task of ASD, indicating that our model is

Table 3 Performance comparison of various methods on ADNI datasets. The best results are bold

Methods	ACC	AUC	Precision	Sensitivity	F1
NCF	0.849±0.018	0.858±0.019	0.926±0.018	0.835±0.029	0.868±0.018
GCN	0.765±0.015	0.618±0.028	0.407±0.195	0.300±0.116	0.321±0.114
f-GCN	0.716±0.009	0.630±0.021	0.779±0.012	0.850±0.016	0.801±0.005
HI-GCN	0.726±0.012	0.695±0.020	0.818±0.014	0.792±0.020	0.793±0.008
T-HI-GCN	0.743±0.008	0.665±0.022	0.817±0.013	0.825±0.013	0.812±0.006
E-HI-GCN	0.814±0.011	0.787±0.019	0.876±0.014	0.879±0.016	0.865±0.007
TE-HI-GCN	**0.894±0.004**	**0.893±0.006**	**0.940±0.005**	**0.912±0.006**	**0.922±0.002**

more effective on the task with limited data. Fine-tuning from the pre-training model helps quickly generate an initial model weight of E-HI-GCN models, thus reducing the requirements of a large labeled training set and accelerating the training stage. Moreover, transfer learning also helps improve the generalization capability of E-HI-GCN in the task of ASD diagnosis, even if the source domain includes little data.

With the ensemble transfer learning scheme on multiple level sparsity networks, TE-HI-GCN achieves improvements of 4.08% and 9.83% over E-HI-GCN on ASD and AD. Transfer learning is not always effective for GCN models. We can see that T-HI-GCN decreases the performance by 0.89% and 0.75% compared with HI-GCN in terms of ACC and AUC. The results confirm that our ensemble transfer strategy with multi-scale networks avoids negative transfer across datasets and achieves the best performance. It also demonstrates that brain networks construction is critical for transfer learning on brain networks data. Transfer learning is more effective when the source and target domains are sparse networks. The ensemble strategy can help achieve transferable representations based on a network topology with multiple levels of sparsity.

Observation (5): In Table 2, the worst performance on the ABIDE is observed for NCF with an average ACC of 0.586 and an AUC of 0.583. The performance of all deep learning methods is better than that of NCF in terms of ACC. Overall, the results suggest that deep learning methods can enable better representation of brain networks as compared to more traditional machine learning methods. However, all the deep learning models except TE-HI-GCN are worse than NCF on the ADNI dataset in Table 3. The reason is that training GCN models from scratch with limited data tend to overfit because we need to optimize a large number of parameters. Compared with GCN models, the feature-based method with brain connectivity requires less training data. It confirms the trend observed: transferring the pre-trained weights is especially advantageous when only limited data is available.

Multiple measurement methods and metrics have been employed in the literature to estimate the connectivity patterns of brain networks from rs-fMRI. The partial and full correlation methods are two dominant approaches to estimate the connectome in rs-fMRI. A partial correlation calculates the interaction between two brain regions after factoring out the contribution to the pairwise correlation that may be due to global effects. Partial correlations relate to the off-diagonal entries of the Inverse Covariance (IC) matrix of the data. Estimation of partial correlations is usually achieved by Maximum Likelihood Estimation (MLE) of the IC matrix. Recently, partial correlation of brain networks obtained through covariance matrix followed by spatial filters was applied to analyze the brain connections. In our study, we employed a partial correlation method with GLASSO (Sparse inverse covariance) to estimate the FC in diagnostic groups and compared their performance with the full correlation method (PCC). A detailed comparison of different correlations in our study is shown in Fig. 5. We found that full correlation features with PCC perform better than partial correlation (GLASSO) approaches regardless of any classification models.

Experiment on Other Parcellations

The parcellations of brain networks can be divided into two categories: predefined structural parcellation atlases and functional atlases. To test whether these observations were dependent on the choice of the atlas, we applied the same methodology on different atlas. We evaluated our model E-HI-GCN on other structural atlases besides AAL: Talariach Daemon (TT) atlas (derived from myeloarchitectonic segmentations) , Harvard-Oxford (HO) atlas (derived from anatomical landmarks: sulci and gyral), Eickhoff-Zilles (EZ) atlas (derived from cytoarchitectonic segmentations). Recently, some functional atlases have been proposed. We also test our model on two functional atlas: CC200 (200 functionally homogeneous regions generated using spatially constrained spectral clustering algorithm), and DOS160 atlas (161-region atlas generated based on meta analysis of task-related fMRI data). From Fig. 6, we find that the result obtained from CC200 works best, which implies that the pairwise correlations among CC200 regions contain more discriminatory patterns than AAL and other atlases. Moreover, we compared some benchmark methods including traditional methods (Ridge, SVM and FCN) and deep learning methods (BrainNet (Kawahara et al., 2017), 3D-CNN (Khosla et al., 2019), ASD-DiagNet (Eslami et al., 2019)) on the multiple atlas in Fig. 7. BrainNet extends convolutional neural networks (CNNs) to handle graph-structured data. Specifically, the edge-to-edge, edge-to-node and node-to-graph convolutional layers are developed to capture topological relationships between network edges. Different from BrainNet that works directly with an adjacency matrix derived from the connectome data, 3D-CNN exploits the 3D spatial structure of rs-fMRI. ASD-DiagNet is a joint learning method combining an autoencoder with a single layer perceptron (SLP) to improve the quality of extracted features and optimized parameters for the classification model. For the traditional classifiers, functional connectivity estimates between pairs of ROIs are vectorized as input. It is apparent that our proposed E-HI-GCN observes improvements upon the previous methods on all the atlases in terms of ACC.

With the idea in (Khosla et al., 2019), we explored another ensemble learning strategy named multi-atlas (MA) ensemble in our work. We chose a variety of so-called atlases, which define a specific parcellation of the brain into ROIs. Each atlas consisted of between 97 and 200 ROIs.

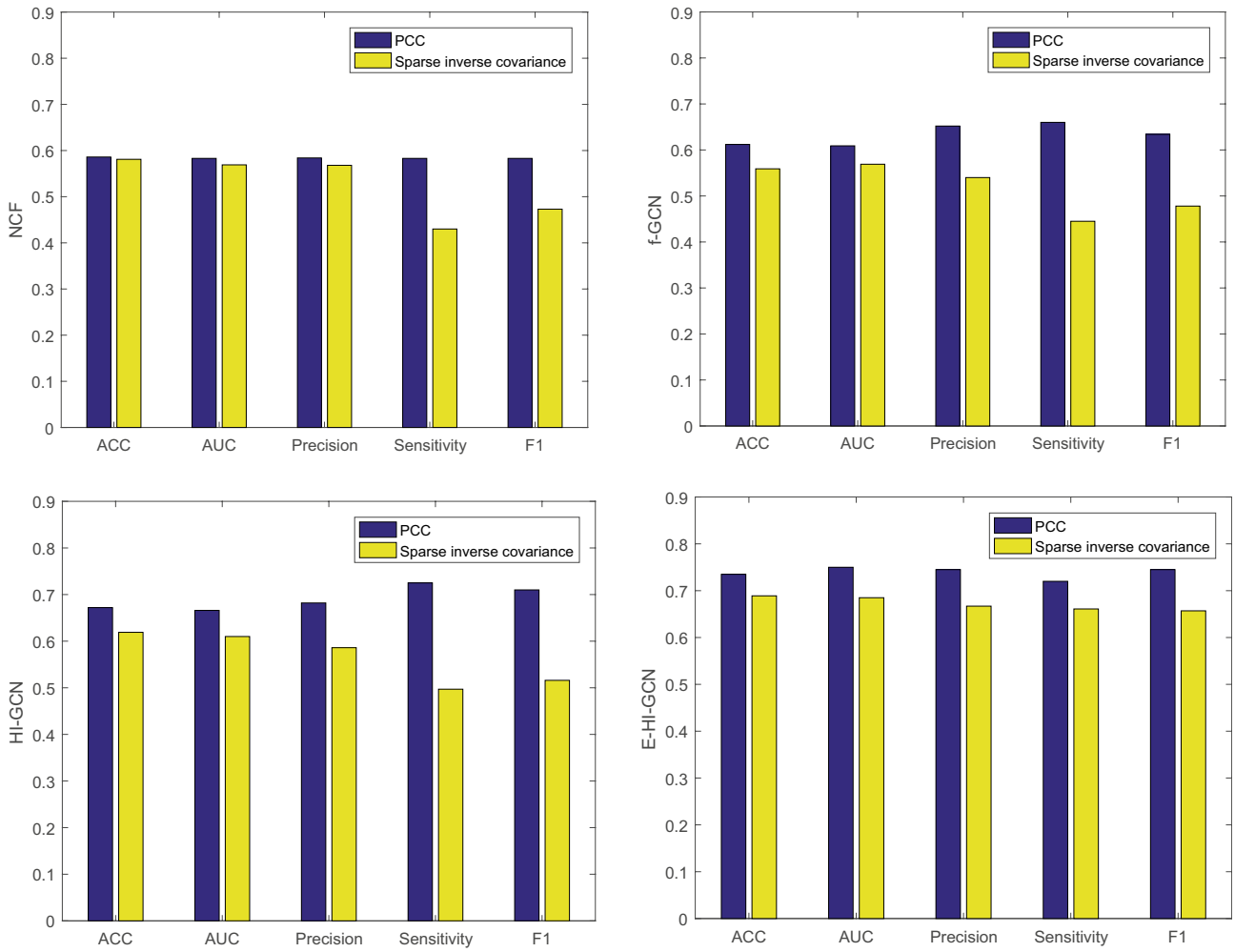

Fig. 5 The comparison of different connectivity (PCC and GLASSO) estimation for NCF, f-GCN, HI-GCN and E-HI-GCN with respect to multiple metrics

Our proposed HI-GCN with the MAEnsemble averages the predictions of the models of different specific methods trained on multiple atlases. For classification, the final prediction is computed as the majority vote of the individual binary class predictions. We applied the multi-atlas (MA) ensemble strategy on the different base classifiers, including Ridge, SVM, FCN, BrainNet, 3D-CNN and HI-GCN. From Fig. 8, we observe that our HI-GCN ensemble obtains better

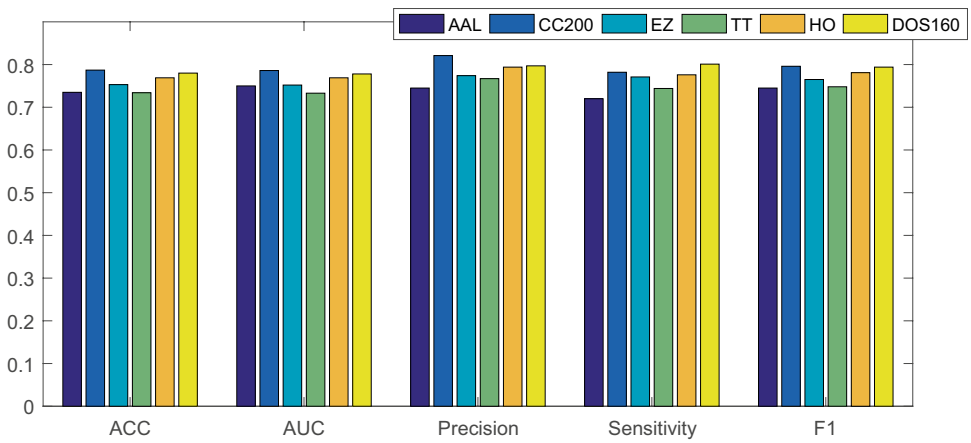

Fig. 6 Performance of E-HI-GCN on multiple atlases

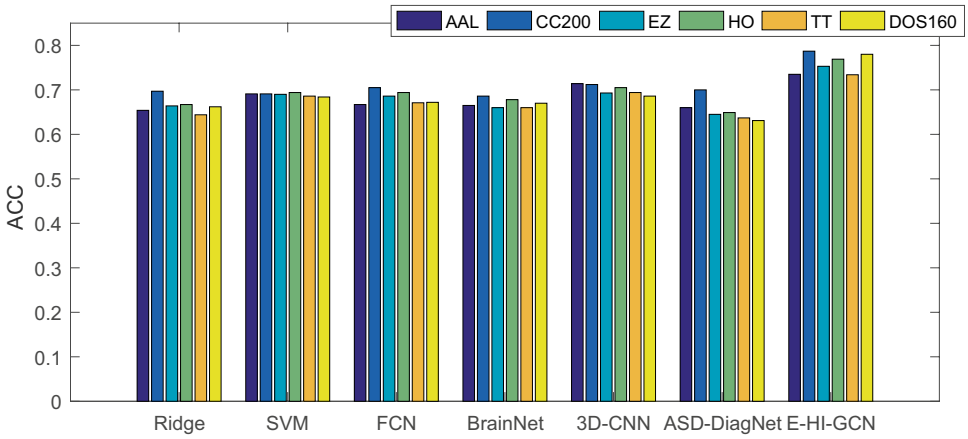

Fig. 7 Performance comparison of different methods on multiple atlases

performance than the other base classifier. The result demonstrated the flexibility advantage of our E-HI-GCN combined with the strength of multiple atlas.

Experiment on HCP Dataset

We apply our proposed E-HI-GCN to predict the gender of healthy individuals of high quality publicly available rs-fMRI dataset: Human Connectome Project (HCP, N=1,096). The Human Connectome Project (HCP) S1200 (Van Essen et al., 2013) contains the rsfMRI data for 1096 young adults (ages 22-35). We used the first session (15 min, 1200 frames, TR=0.72s) for each subject and excluded 5 rs-fMRIs with less than 1200 frames, resulting in data of 498 females and 593 males. Each rs-fMRI went through the minimal processing pipeline of HCP, fMRISurface (Glasser et al., 2016), which mapped each volume time series to the standard CIFTI grayordinates space. The cortical surface was parcellated to 22 major ROIs (Glasser et al., 2013), and the average BOLD signal in each ROI was normalized to z-scores.

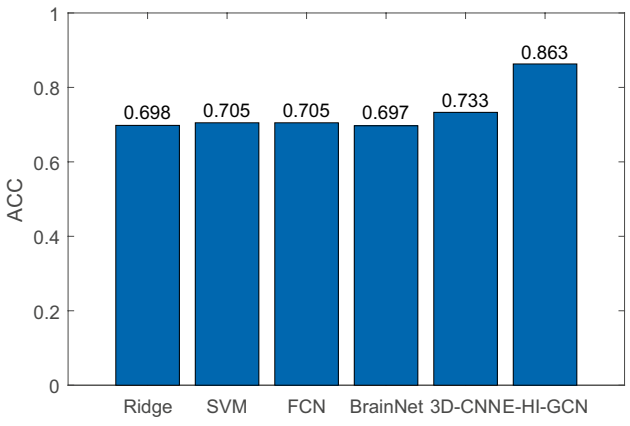

Fig. 8 Performance comparison of the multi-atlas (MA) ensemble strategy with different base classifier methods

As we show in Table 4, the proposed HI-GCN and E-HI-GCN bring improvements to the prediction performance. It can be verified that removing the high dimensional noisy connections and modeling the correlation among the instances contributes to the performance improvements. Moreover, the ensemble learning strategy for combining predictions from multi-scale networks helps promote the learning performance.

Interpretability

Each item $f_{i,j}$ in $\mathcal{F}$ can be interpreted as the membership of the node i to the cluster S_j. With the optimized $\mathcal{F}$, the function $m(\cdot)$ is the cluster mapping function which maps the nodes to the corresponding clusters according to the membership optimized. The score of the p-th subnetwork SN_p is calculated as:

$$Score_p = \frac{1}{n_p^2} \sum_{i,j \in SN_p \text{and } m(i) \neq m(j)} f_{i,m(i)} f_{j,m(j)} \tag{9}$$

where n_p is the amount of the brain regions of p-th subnetwork.

Moreover, the inter-network connections in this system play essential roles. To explore the important cross-subnetwork correlations, the correlation score of two subnetworks SN_p and SN_q is calculated as:

$$CorrScore_{p,q} = \frac{1}{n_p^2 n_q^2} \sum_{i \in SN_p, j \in SN_q \text{and } m(i) \neq m(j)} f_{i,m(i)} f_{j,m(j)} \tag{10}$$

where n_p and n_q are the amounts of the brain region of subnetworks.

In neuroscience studies, researchers are not only interested in providing a better prediction model but also in identifying which brain areas are more affected by the disease. This can help to diagnose the early stages of the disease and how it spreads. Rs-fMRI studies in neurotypical

Table 4 Performance comparison of various methods on HCP dataset. The best results are bold

Methods	ACC	AUC	Precision	Sensitivity	F1
GCN	0.687±0.002	0.679±0.002	0.691±0.003	0.738±0.005	0.710±0.003
f-GCN	0.595±0.002	0.581±0.002	0.612±0.002	0.736±0.025	0.657±0.004
HI-GCN	0.736±0.002	0.733±0.003	0.765±0.005	**0.770±0.006**	0.761±0.001
E-HI-GCN	**0.750±0.002**	**0.750±0.002**	**0.783±0.003**	0.758±0.003	**0.767±0.001**

individuals have identified several major intrinsically connected networks related to visual, motor, auditory, memory and executive processes. It is envisaged that the identification of brain networks using FC may provide potential biomarkers to understand the organization and alterations of brain networks in ASD. To better understand the cortical circuitry underlying the connectivity between large-scale neural networks, we conduct our model with data-driven to investigate potential intrinsically within-network and between-network connectivity. One of the strengths of our E-HI-GCN is that it facilitates the identification of biomarkers due to the node clustering property of our f-GCN. In this experiment, we empirically investigate the effectiveness of subnetwork and inter-subnetwork connections identification. We evaluate eight networks including DMN (default mode network), DAN (dorsal attention), AN (auditory network), CN (core network), SN (salience network), SMN (somato-motor network), VN (visual network), CEN (central executive network) with brain networks templates provided in (Mantini et al., 2009). Top 3 subnetworks and the top 5 inter-subnetworks selected by our E-HI-GCN are shown in Table 5, Figs. 9 and 10.

We found that the top 3 subnetworks identified by E-HI-GCN yielded promising patterns expected from prior knowledge on neuroimaging and cognition. These included the CEN, DMN and SN, which are three core neurocognitive networks. Neuroimaging studies have demonstrated that ASD is associated with the altered functional connectivity of the three neurocognitive networks that are hypothesized to be central to the symptomatology of ASD. The default mode network (DMN) consists of key nodes in the posterior cingulate cortex (PCC), the medial temporal lobes (MTL) and the medial prefrontal cortex (MPFC) and is active in self-related tasks such as autobiographical memories or social tasks such as the theory of mind. The salience network (SN) involves the anterior insula (AI) and the anterior cingulate cortex (ACC) and is thought to regulate the switching of endogenous and exogenous attention to relevant stimuli which help in guiding behavior.

Table 5 Top 3 intra-subnetworks and the top 5 cross inter-subnetworks selected and weights optimized by our model

Subnetworks		Correlation between subnetworks	
subnetwork name	weight	inter-subnetworks name	weight
CEN	0.046	(CEN,SMN)	0.078
DMN	0.040	(CEN,SN)	0.075
SN	0.039	(DMN,SMN)	0.062
-	-	(DMN, CEN)	0.061
-	-	(DMN,VN)	0.059

We also investigated the cross-network interactions. The inter-subnetwork pairs included (CEN,SMN), (CEN,SN), (DMN,SMN), (DMN,CEN) and (DMN,VN). The findings in our study are consistent with results reported in previous studies. Neuroimaging studies have demonstrated that dysfunction of triple networks including CEN, SN, and DMN were associated with ASD. Notably, the SN, CEN, and DMN are often co-activated or deactivated during attentionally-demanding tasks, suggesting that these networks function in concert to support attention and cognition. In particular, the triple-network model posits a central role for the SN in initiating switching between the CEN and DMN, a process essential for attention and flexible cognitive control. Moreover, from a clinical standpoint, anticorrelated contributions from regions of the DMN and SMN have been previously reported in ASD (Lynch et al., 2013) (Nebel et al., 2016). Besides, some between networks FCs including DMN-VN keep their sign of relation to diagnosis variable at different frequency bands diagnosing ASD (Lee & Frangou, 2017).

Ablation Study and Discussion

The Effectiveness of Transfer Learning

Pre-training is crucial for learning deep neural networks. The fundamental problems that need to be considered for reliability is what to transfer. There are several options for pre-training, as shown in Fig. 4. In this study, We conduct an ablation study in Tables 6 and 7 to explore several transfer learning strategies to investigate which one is best to leverage the source model and adapt it to the target task, including node clustering pretraining (MGC module in f-GCN) pretraining, node-level feature learning (GCN module in f-GCN) pretraining, the whole f-GCN pretraining, graph-level embedding learning (p-GCN) pretraining.We conduct the systematic investigation of pre-training strategies for our GCN model from the two aspects:

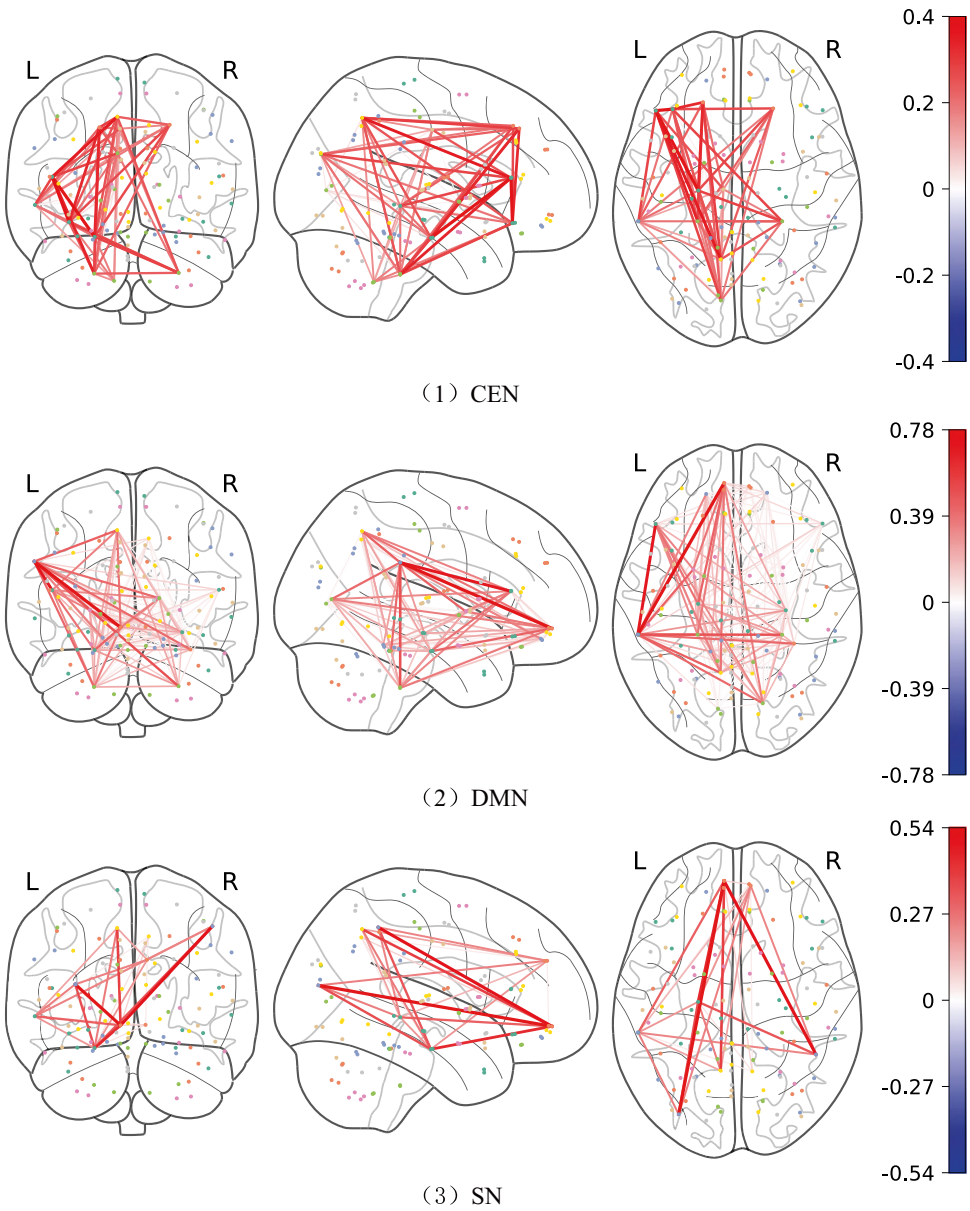

Fig. 9 The top 3 subnetworks identified by our E-HI-GCN

1. ASD → AD Except transfer on the GCN module of f-GCN, all the transfers strategies are effective and improve the performance of target models of AD. This transfer learning strategy is able to enhance GCN classification when the target dataset has a limited sample size. Our findings show that the best strategy is pre-training the whole f-GCN. The conclusion can be drawn that pre-training f-GCN can capture generic structural features in brain networks for two diseases and improve the performance of target models. The network structural characteristics among the brain regions are shared across the diseases. The pre-training on GCN module of f-GCN performs worse, which suggests that transferring the GCN module of f-GCN results in negative transfer. This is because the task of node embedding learning in brain networks for ASD might be unrelated to the task for AD and can even hurt the downstream performance (negative transfer).
2. AD → ASD From Table 7, it is surprising that only the pre-training p-GCN is effective. The reason may be that the limited data in AD cannot train a generalized model, the learned parameters of which are not appropriate the ASD. The parameters of p-GCN module are relatively fewer. Hence, the p-GCN is easy to be trained well on the AD with limited data, so that graph embeddings are robust and transferable across the domains. The learning of network embeddings with considering the sample correlation can generalize across tasks. Thus it can be well transferred to the ASD. Although it actually further confirms our finding that the two diseases are correlated,

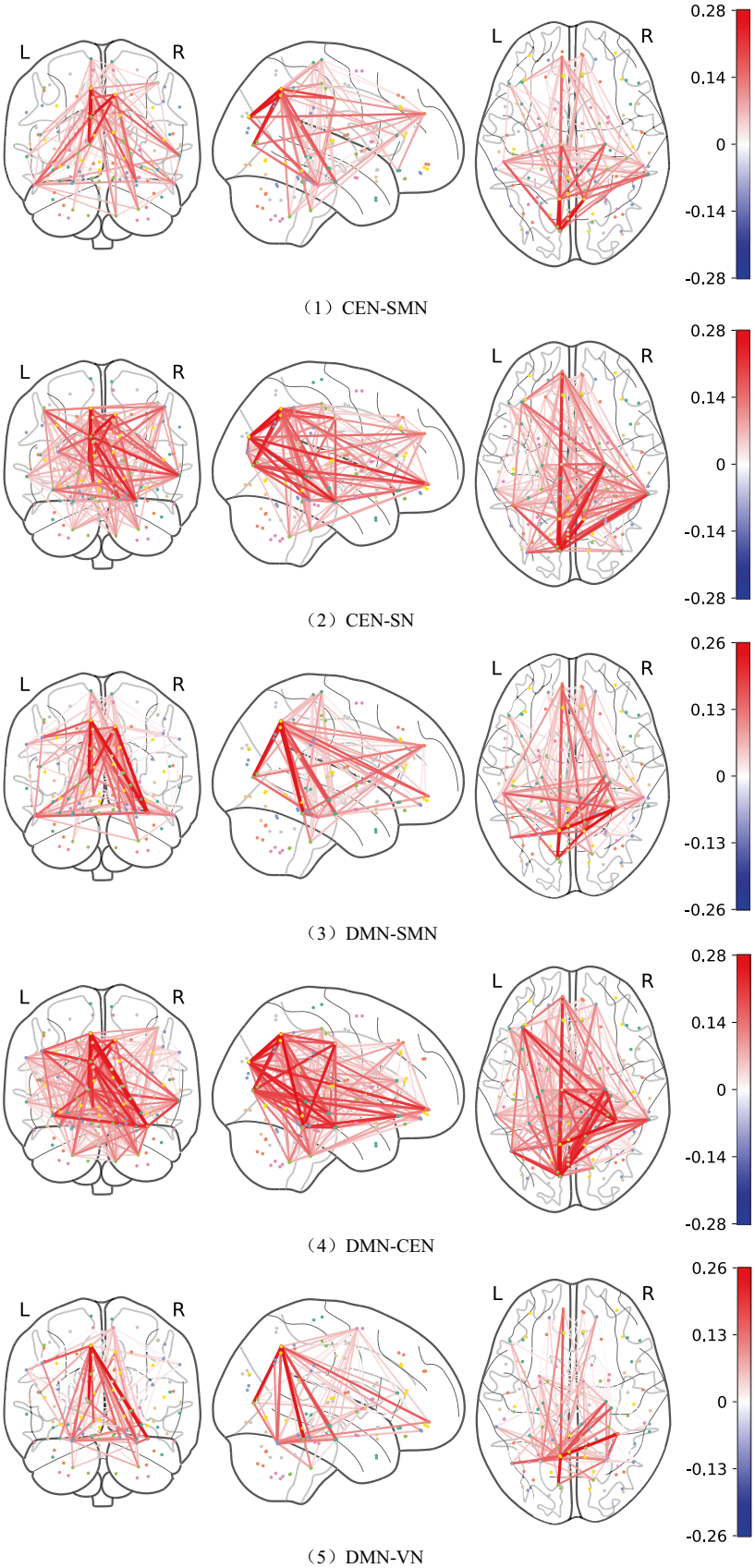

Fig. 10 The top 5 inter-subnetworks identified by our E-HI-GCN

Table 6 Performance of different pre-training strategies on ABIDE dataset.The best results are bold

Pre-learning	ACC	AUC	Precision	Sensitivity	F1
f-GCN	0.542±0.002	0.575±0.003	0.557±0.008	0.529±0.018	0.523±0.08
MGC in f-GCN	0.542±0.002	0.546±0.002	0.500±0.058	0.620±0.005	0.504±0.025
GCN in f-GCN	0.700±0.001	0.699±0.001	0.679±0.003	0.674±0.003	0.678±0.001
p-GCN	**0.765±0.003**	**0.762±0.003**	**0.779±0.005**	**0.799±0.009**	**0.784±0.003**
E-HI-GCN	0.583±0.001	0.572±0.001	0.587±0.005	0.448±0.019	0.476±0.006

the learned knowledge for transferring is different for ASD → AD and AD → ASD. An appropriate transfer learning for learned knowledge from the source dataset can help solve the target task.

The Effectiveness of Ensemble Learning

Finally, we examined the behavior of the ensemble method by means of varying a different numbers of threshold values, which plays a crucial role in the ensemble performance. The effectiveness of ensemble with respect to the ensemble capacity is as shown in Fig. 11 and Table 8. As can be seen that more component classifiers with different numbers of threshold values can achieve better performance. The best performance can be achieved by the ensemble method with 15 classifiers in terms of ACC (0.825) and AUC (0.823), respectively.

Comparisons with Prior Works

Table 9 shows the comparable results of our ASD vs. NC classification with prior works in terms of accuracy as reported in the respective references. In general, two types of methods are usually developed for the diagnosis of mental disorder diseases: (1) the traditional machine learning procedure (feature extraction and classification) and (2) the deep learning procedure (an end-to-end procedure). For the traditional machine learning procedure, a straightforward solution that has been extensively explored is to first derive features from brain networks. Dadi et al. (2019) conducted a sufficient comparison (8 different ways of defining regions-either pre-defined or generated from the rest-fMRI data, (3) measures to build functional connectomes from the extracted time-series, and 10 classification models to compare functional interactions across subjects). Through the comparison, the optimal choices in functional connectivity prediction pipeline brain regions defined with regions using DictLearn, connectivity matrices parametrized by their tangent-space representation, and an l2-regularized logistic regression as a classier. On the ABIDE dataset, the best accuracy is 71.1% (median) and 75.6% (the 95th percentile). Abraham also employed the same strategy with the best pipeline and obtained a similar performance (accuracy is 66.8%) (Abraham et al., 2017). The most important limitation of the traditional machine learning procedure is that feature extraction and model learning are treated as two separate tasks in these methods, so potential inconsistency between human-engineered features and classifiers may degrade the final performance of these methods. Moreover, relying solely on subject-specific imaging feature vectors fails to model the interaction and similarity between subjects, which can reduce performance. Compared with the traditional machine learning procedure, the expressive power of deep learning to extract the underlying complex patterns from data has been well recognized. The power of deep learning lies in automatically learning relevant and powerful features for any perdition task, which is made possible through end-to-end architectures. The result demonstrates that the deep learning based end-to-end methods achieve a better performance in network-structured learning tasks. Heinsfeld et al. (2018) and Eslami et al. (2019) have proposed a joint learning procedure using an autoencoder and a single layer or multi-layer perceptron, which results in improved quality of extracted features and optimized parameters for the model. To the best of our knowledge, the work of Parisot et al. (2017) is currently relevant with ours for ASD diagnosis on the whole ABIDE dataset. Parisot et al. define a subject's feature vector as its vectorised functional

Table 7 Performance of different pre-training strategies on ADNI dataset.The best results are bold

Pre-learning	ACC	AUC	Precision	Sensitivity	F1
f-GCN	**0.894±0.004**	**0.893±0.006**	0.940±0.005	**0.912±0.060**	**0.922±0.002**
MGC in f-GCN	0.890±0.006	0.890±0.007	0.939±0.005	0.907±0.009	0.918±0.004
GCN in f-GCN	0.797±0.011	0.745±0.021	0.839±0.015	0.907±0.014	0.858±0.006
p-GCN	0.821±0.009	0.798±0.016	0.877±0.012	0.884±0.011	0.871±0.005
E-HI-GCN	0.849±0.008	0.857±0.012	**0.943±0.008**	0.846±0.021	0.880±0.007

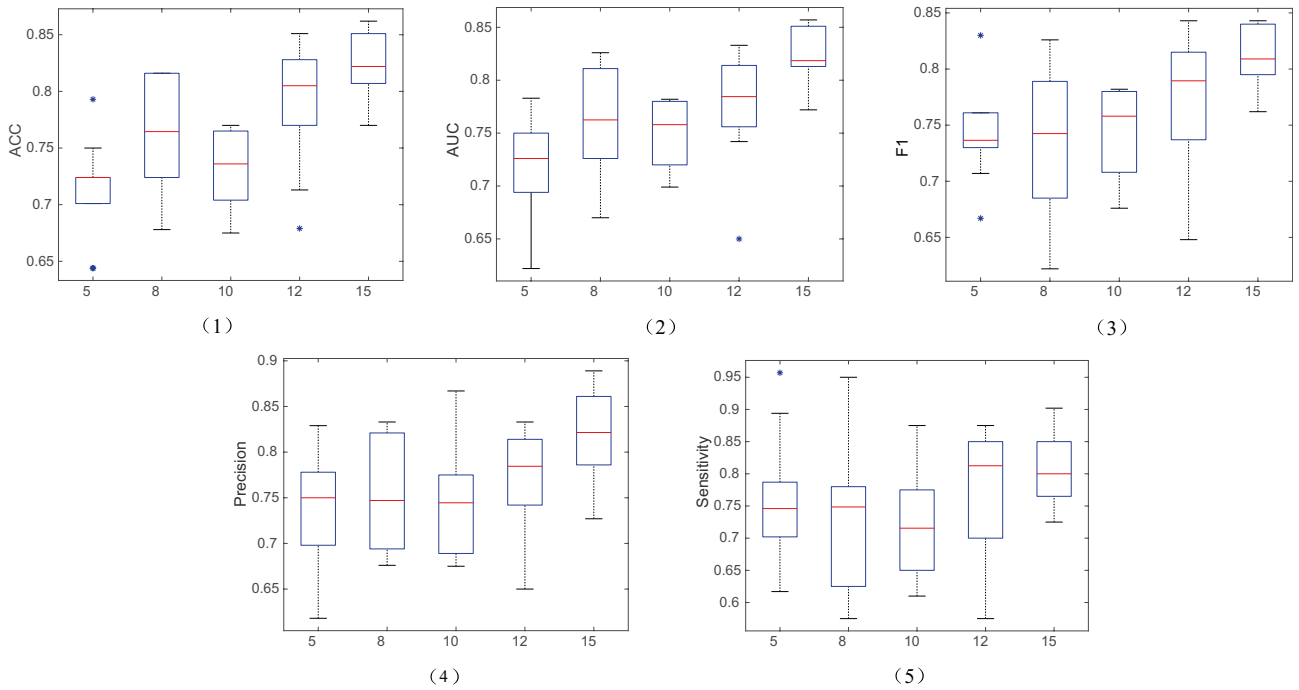

Fig. 11 The results of our E-HI-GCN with different numbers of threshold values

connectivity matrix and employ a ridge classifier to select the most discriminative features from the training set. With the selected connectivity information as the subject's feature, a population is represented as a graph where its vertices are associated with the extracted image-based feature vectors and the edges encode non-imaging measures (gender and acquisition site). The significant difference of our HI-GCN from the model in Parisot et al. (2017) are: 1) **node**: the embedding features of the nodes in the population is learned automatically rather than extracted; 2) **edge**: the similarity between nodes is calculated considering the structure of the brain functional network when constructing the population network; 3) **induction**: the GCN model proposed in Parisot et al. (2017) is a transductive method with a fixed population network. They do not naturally generalize to unseen data since they make predictions on nodes in the fixed population network. Our HI-GCN is an inductive learning method and generalizes to produce embeddings for unseen nodes. Experiments demonstrate that the proposed HI-GCN model performs better than the GCN model proposed in Parisot et al. (2017), indicating simultaneously taking both the brain regions correlations and subject correlations into account is important. The apparent limitation of such a model is that they can only learn on the vectorized node, which cannot effectively generalize the condition that the node is a graph describing the functional connectivity. The graph representation techniques have recently shifted from hand-crafted kernel methods to deep learning based end-to-end methods, which achieve better performance in graph-structured learning tasks. Moreover, the feature extracted prior to the classification may not be appropriate for GCN classification due to lacking the capacity of jointly learning. From Table 10, we can observe that our models usually achieve competitive performance against the state-of-the-art methods. We

Table 8 The results of our E-HI-GCN with different numbers of threshold values. The best results are bold

Number of threshold values	Range of threshold values	ACC	AUC	Precision	Sensitivity	F1
5	[0.1,0.5]	0.717±0.002	0.714±0.002	0.737±0.003	0.758±0.003	0.740±0.002
8	[0.05,0.4]	0.762±0.002	0.760±0.003	0.752±0.003	0.733±0.010	0.736±0.004
10	[0.05,0.5]	0.735±0.001	0.750±0.001	0.745±0.003	0.720±0.007	0.745±0.001
12	[0.05,0.6]	0.788±0.003	0.786±0.003	0.770±0.003	0.774±0.009	0.768±0.004
15	[0.05,0.75]	0.825±0.001	0.823±0.001	0.817±0.003	0.806±0.003	0.809±0.001

Table 9 The comparison among different classifiers with previous methods for ASD vs. NC on ABIDE dataset

Method	Feature	Classifier	Data	Atlas	CV	ACC
.Parisot et al. (2017)	Brain connectivity feature	GCN	403 ASD vs. 468 NC	Harvard Oxford (HO)	10-CV	69.5%
.Abraham et al. (2017)	Tangent space embedding	l2-regularized classifiers	403 ASD vs. 468 NC	data-driven atlas (dictionary learning)	Inter-site 10-CV	66.8%
.Dadi et al. (2019)	Network connectivity feature	l2-regularized logistic regression	402 ASD vs. 464 NC	data-driven atlas (dictionary learning)	random 100-CV	69.7%
.Wong et al. (2018)	Riemannian geometry feature	logistic regression	403 ASD vs. 468 NC	HO	10CV	71.1%
.Eslami et al. (2019)	No feature extraction	ASD-DiagNet (with aug)	505 ASD vs. 530 NC.	CC200	10-CV	69.4% (70.3%)
.Eslami et al. (2019)	No feature extraction	ASD-DiagNet (with aug)	505 ASD vs. 530 NC.	AAL	10-CV	64.5% (67.5%)
.Eslami et al. (2019)	No feature extraction	ASD-DiagNet (with aug)	505 ASD vs. 530 NC.	TT	10-CV	65.2% (65.3%)
.Heinsfeld et al. (2018)	No feature extraction	Two stacked denoising autoencoders	505 ASD vs. 530 NC	CC200	10-CV	70%
.Dvornek et al. (2018)	rs-fMRI time-series+phenotypic features	LSTM	529 ASD vs. 571 NC	CC200	10-CV	70.1%
.Dvornek et al. (2017)	rs-fMRI time-series	LSTM	529 ASD vs. 571 NC	CC200	10-CV	66.8%
.Sherkatghanad et al. (2020)	No feature extraction	CNN	505 ASD vs. 530 NC.	CC400	10-CV	70.2%
.Nielsen et al. (2013)	Network connectivity feature	general linear model	539 ASD vs. 573 NC	no atlas	leave-one-out	60.0%
.Xing et al. (2018)	No feature extraction	CNN with element-wise filters	569 ASD vs. 572 NC	AAL	5-CV	66.8%
.Aghdam et al. (2018)	No feature extraction	Deep belief Network (DBN)	116 ASD vs. 69 NC	AAL	10-CV	65.56%
.Kazeminejad and Sotero (2020)	Graph features	MLP	493 ASD vs. 520 NC	AAL	cross-validation sets	55.50%
.Kazeminejad and Sotero (2020)	PCA + Graph features	MLP	493 ASD vs. 520 NC	CC200	cross-validation sets	62.40%
Our method	No feature extraction	HI-GCN	403 ASD vs. 468 NC	AAL	10-CV	67.2%
Our method	No feature extraction	E-HI-GCN(10)	403 ASD vs. 468 NC	AAL	10-CV	**73.5%**
Our method	No feature extraction	E-HI-GCN(15)	403 ASD vs. 468 NC	AAL	10-CV	**82.5%**
Our method	No feature extraction	TE-HI-GCN	403 ASD vs. 468 NC	AAL	10-CV	**76.5%**
Our method	No feature extraction	E-HI-GCN	403 ASD vs. 468 NC	CC200	10-CV	**78.7%**
Our method	No feature extraction	E-HI-GCN	403 ASD vs. 468 NC	MA ensemble	10-CV	**86.3%**

improved the previously reported results and obtained 76.5% on the AAL atlas, 78.7% on the CC200 atlas, and 86.3% with an ensemble on the multi-atlas in the distinction of ASD from control subjects, respectively. Overall, the results demonstrate that our models can improve upon state-of-the-art algorithms not just in traditional classification methods but also in deep learning methods. Although the experimental setup in these references is slightly different, this table shows that our end-to-end classification model can classify the subjects more accurately.

Table 10 The comparison among different classifiers with previous methods for AD vs. MCI on ADNI dataset

Method	Feature	Classifier	Data	CV	ACC
Dadi et al. (2019)	Network connectivity feature	l2-regularized logistic regression	40 AD vs. 96 MCI	random 100-CV (75% for training, and 25% for test)	72.2%
Our method	No feature extraction	HI-GCN	34 AD vs. 99 MCI	10-CV	83.6%
Our method	No feature extraction	T-HI-GCN	34 AD vs. 99 MCI	10-CV	85.0%
Our method	No feature extraction	E-HI-GCN	34 AD vs. 99 MCI	10-CV	85.1%
Our method	No feature extraction	TE-HI-GCN	34 AD vs. 99 MCI	10-CV	86.7%

We also compare our methods with several recent state-of the-art methods reported in the literature using rs-fMRI data for AD vs. MCI classification. Few works for discriminating AD and MCI with functional brain networks exist in the literature. The only comparable method is proposed in Dadi et al. (2019). The best accuracy is 72.5% (median) and 84.5% (the 95th percentile) over cross-validation folds (n = 100) on the ADNI dataset with the same samples. Results obtained for the ADNI database show an increase in performance with respect to the competing method in Table 10.

Conclusion

Recently, functional connectivity networks constructed from the rs-fMRI are holding great promise for distinguishing the disorder patients from NC. Network embedding is aimed at learning compact node representations based on network topology to facilitate the task of network classification. Deep learning models have an enormous capacity of network embedding with a huge number of parameters that need to be trained during the learning process. The rs-fMRI data is in general of high dimension with a small sample size, and the use of deep learning in small data sets still remains a big challenge. In order to achieve a better network embedding from brain networks, we propose a hierarchical GCN framework for modeling the brain connectivity network and population network simultaneously to learn the network feature embedding with considering the network topology information and subject's association. Moreover, we develop a transfer learning scheme enabling GCN to learn generic graph structural features by leveraging the commonality in two related domains. To transfer the appropriate knowledge for the network data avoiding negative transferring, the transfer learning is also carefully conducted on the multiple levels of topological structure in the original connectivity network. Extensive experiments are conducted on two real medical clinical applications: diagnosis of ASD and diagnosis of Alzheimer's disease, which demonstrates network embedding learning from exploring the data correlations and transferring from the related domains can improve prediction performance. Moreover, we also explore the hypothesis that the diagnosis model of brain disorders with GCN can be transferred across relevant diseases. This is in line with the conclusion drawn from previous medical image-based computer-aided detection studies that transfer learning could be a useful technique to mitigate the issue due to a small well-annotated dataset in the medical imaging domain. Furthermore, our study found that some connections within subnetworks and between subnetworks have a more significant role in the classification of diagnostic groups.

Information Sharing Statement

The data sets used in this study are freely available and can be downloaded from the corresponding project websites- http://fcon.1000.projects.nitrc.org/indi/abide/ for ABIDE dataset and http://adni.loni.usc.edu/ for ADNI dataset and the corresponding descriptions are shown in Sec. 5.1 Databases and Preprocessing. The code for the TE-HI-GCN is available at: https://github.com/llt1836/TE-HI-GCN.

Acknowledgements This research was supported by the National Natural Science Foundation of China (No.62076059) and the Fundamental Research Funds for the Central Universities (No. N2016001).

References

A Khan, S., A Khan, S., R Narendra, A., Mushtaq, G., A Zahran, S., Khan, S., & A Kamal, M. (2016). Alzheimer's disease and autistic spectrum disorder: Is there any association? *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders) 15*(4), 390–402.

Abraham, A., Milham, M. P., Di Martino, A., Craddock, R. C., Samaras, D., Thirion, B., & Varoquaux, G. (2017). Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage, 147,* 736–745.

Aghdam, M. A., Sharifi, A., & Pedram, M. M. (2018). Combination of rs-fmri and smri data to discriminate autism spectrum disorders in young children using deep belief network. *Journal of Digital Imaging, 31*(6), 895–903.

Anirudh, R., & Thiagarajan, J. J. (2019). Bootstrapping graph convolutional neural networks for autism spectrum disorder classification. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 3197–3201.

Arslan, S., Ktena, S. I., Glocker, B., & Rueckert, D. (2018). Graph saliency maps through spectral convolutional networks: Application to sex classification with brain connectivity. In *Graphs in Biomedical Image Analysis and Integrating Medical Imaging and Non-Imaging Modalities*. Springer, pp. 3–13.

Bajestani, G. S., Behrooz, M., Khani, A. G., Nouri-Baygi, M., & Mollaei, A. (2019). Diagnosis of autism spectrum disorder based on complex network features. *Computer Methods and Programs in Biomedicine, 177,* 277–283.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (compcor) for bold and perfusion based fmri. *Neuroimage, 37*(1), 90–101.

Betzel, R. F., & Bassett, D. S. (2017). Multi-scale brain networks. *Neuroimage, 160,* 73–83.

Chen, X., Zhang, H., Lee, S. -W., & Shen, D. (2017). Hierarchical high-order functional connectivity networks and selective feature fusion for mci classification. *Neuroinformatics, 15*(3), 271–284.

Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., Khundrakpam, B. S., Lewis, J. D., Li, Q., Milham, M., et al. (2013). The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics,* 7.

Dadi, K., Rahim, M., Abraham, A., Chyzhyk, D., Milham, M., Thirion, B., et al. (2019). Benchmarking functional connectome-based predictive models for resting-state fmri. *NeuroImage, 192,* 115–134.

Di Martino, A., Yan, C. -G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry, 19*(6), 659–667.

Duc, N. T., Ryu, S., Qureshi, M. N. I., Choi, M., Lee, K. H., & Lee, B. (2020). 3d-deep learning based automatic diagnosis of alzheimer's disease with joint mmse prediction using resting-state fmri. *Neuroinformatics, 18*(1), 71–86.

Dvornek, N. C., Ventola, P., & Duncan, J. S. (2018). Combining phenotypic and resting-state fmri data for autism classification with recurrent neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, pp. 725–728.

Dvornek, N. C., Ventola, P., Pelphrey, K. A., & Duncan, J. S. (2017). Identifying autism from resting-state fmri using long short-term memory networks. In *International Workshop on Machine Learning in Medical Imaging*, Springer, pp. 362–370.

Ebrahimighahnavieh, M. A., Luo, S., & Chiong, R. (2020). Deep learning to detect alzheimer's disease from neuroimaging: A systematic literature review. *Computer Methods and Programs in Biomedicine, 187,* 105242.

Eid, O. M., & Eid, M. M. (2019). The implications of genetic factors in autism spectrum disorder and alzheimer's disease. *Neurological Disorders and Imaging Physics*.

Eslami, T., Mirjalili, V., Fong, A., Laird, A. R., & Saeed, F. (2019). Asd-diagnet: a hybrid learning approach for detection of autism spectrum disorder using fmri data. *Frontiers in Neuroinformatics, 13,* 70.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102*(27), 9673–9678.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. -P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping, 2*(4), 189–210.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature, 536*(7615), 171–178.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage, 80,* 105–124.

Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge discovery and Data Mining*, pp. 855–864.

Guo, H., Liu, L., Chen, J., Xu, Y., & Jie, X. (2017). Alzheimer classification using a minimum spanning tree of high-order functional network on fmri dataset. *Frontiers in Neuroscience, 11,* 639.

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: Clinical 17*, 16–23.

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

Kazeminejad, A., & Sotero, R. C. (2020). The importance of anti-correlations in graph theory based classification of autism spectrum disorder. *Frontiers in Neuroscience, 14,* 676.

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2016). Application of advanced machine learning methods on resting-state fmri network for identification of mild cognitive impairment and alzheimer's disease. *Brain Imaging and Behavior 10*(3), 799–817.

Khosla, M., Jamison, K., Kuceyeski, A., & Sabuncu, M. R. (2019). Ensemble learning with 3d convolutional neural networks for functional connectome-based prediction. *NeuroImage, 199,* 651–662.

Khosla, M., Jamison, K., Ngo, G. H., Kuceyeski, A., & Sabuncu, M. R. (2019). Machine learning in resting-state fmri analysis. *Magnetic Resonance Imaging, 64,* 101–121.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

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, W. H., & Frangou, S. (2017). Linking functional connectivity and dynamic properties of resting-state networks. *Scientific Reports, 7*(1), 1–10.

Li, G., Muller, M., Thabet, A., & Ghanem, B. (2019). Deepgcns: Can gcns go as deep as cnns? In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9267–9276.

Li, Q., Han, Z., & Wu, X. -M. (2018). Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 32.

Li, X., & Duncan, J. (2020). *Braingnn: Interpretable brain graph neural network for fmri analysis. bioRxiv*.

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*, Springer, pp. 485–493.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42,* 60–88.

Lund, T. E., Nørgaard, M. D., Rostrup, E., Rowe, J. B., & Paulson, O. B. (2005). Motion or activity: their role in intra-and inter-subject variation in fmri. *Neuroimage, 26*(3), 960–964.

Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik, 29*(2), 102–127.

Lynch, C. J., Uddin, L. Q., Supekar, K., Khouzam, A., Phillips, J., & Menon, V. (2013). Default mode network in childhood autism: posteromedial cortex heterogeneity and relationship with social deficits. *Biological Psychiatry, 74*(3), 212–219.

Ma, Y., Wang, S., Aggarwal, C. C., & Tang, J. (2019). Graph convolutional networks with eigenpooling. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 723–731.

Mantini, D., Corbetta, M., Perrucci, M. G., Romani, G. L., & Del Gratta, C. (2009). Large-scale brain networks account for sustained and transient activity during target detection. *Neuroimage, 44*(1), 265–274.

Mier, W., & Mier, D. (2015). Advantages in functional imaging of the brain. *Frontiers in Human Neuroscience, 9,* 249.

Nasrat, A. M., Nasrat, R. M., & Nasrat, M. M. (2017). Autism and alzheimer; the etiopathologic twins. *American Journal of Medicine and Medical Sciences*.

Nebel, M. B., Eloyan, A., Nettles, C. A., Sweeney, K. L., Ament, K., Ward, R. E., et al. (2016). Intrinsic visual-motor synchrony correlates with social deficits in autism. *Biological Psychiatry, 79*(8), 633–641.

Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., et al. (2013). Multisite functional connectivity mri classification of autism: Abide results. *Frontiers in Human Neuroscience, 7,* 599.

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., & Rueckert, D. (2018). 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., Moreno, R. G., Glocker, B., & Rueckert, D. (2017). Spectral graph convolutions for population-based disease prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 177–185.

Qi, S., Meesters, S., Nicolay, K., ter Haar Romeny, B. M., & Ossenblok, P. (2015). The influence of construction methodology on structural brain network measures: A review. *Journal of Neuroscience Methods, 253,* 170–182.

Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. arXiv preprint arXiv:1902.07208.

Sherkatghanad, Z., Akhondzadeh, M., Salari, S., Zomorodi-Moghadam, M., Abdar, M., Acharya, U. R., et al. (2020). Automated detection of autism spectrum disorder using a convolutional neural network. *Frontiers in Neuroscience, 13,* 1325.

Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015). Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 1067–1077.

Tang, W., Lu, Z., & Dhillon, I. S. (2009). Clustering with multiple graphs. In *2009 Ninth IEEE International Conference on Data Mining*, IEEE, pp. 1016–1021.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage, 15*(1), 273–289.

Vaishali, S., Rao, K. K., & Rao, G. S. (2015). A review on noise reduction methods for brain mri images. In *2015 International Conference on Signal Processing and Communication Engineering Systems*, IEEE, pp. 363–365.

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

Wang, J., Zuo, X., & He, Y. (2010). Graph-based network analysis of resting-state functional mri. *Frontiers in Systems Neuroscience, 4,* 16.

Wang, M., Hao, X., Huang, J., Wang, K., Shen, L., Xu, X., et al. (2020). Hierarchical structured sparse learning for schizophrenia identification. *Neuroinformatics, 18*(1), 43–57.

Wang, X., Zhen, X., Li, Q., Shen, D., & Huang, H. (2018). Cognitive assessment prediction in alzheimer's disease by multi-layer multi-target regression. *Neuroinformatics, 16*(3), 285–294.

Wong, E., Anderson, J. S., Zielinski, B. A., & Fletcher, P. T. (2018). Riemannian regression and classification models of brain networks applied to autism. In *International Workshop on Connectomics in Neuroimaging*, Springer, pp. 78–87.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*.

Xing, X., Ji, J., & Yao, Y. (2018). Convolutional neural network with element-wise filters to extract hierarchical topological features for brain networks. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, pp. 780–783.

Yao, D., Liu, M., Wang, M., Lian, C., Wei, J., Sun, L., Sui, J., & Shen, D. (2019). Triplet graph convolutional network for multi-scale analysis of functional connectivity using functional mri. In *International Workshop on Graph Learning in Medical Imaging*, Springer, pp. 70–78.

Yue, X., Wang, Z., Huang, J., Parthasarathy, S., Moosavinasab, S., Huang, Y., et al. (2020). Graph embedding on biomedical networks: methods, applications and evaluations. *Bioinformatics, 36*(4), 1241–1251.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020). Graph neural networks: A review of methods and applications. *AI Open, 1,* 57–81.

Zhou, Z., Sodha, V., Siddiquee, M. M. R., Feng, R., Tajbakhsh, N., Gotway, M. B., & Liang, J. (2019). Models genesis: Generic autodidactic models for 3d medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 384–393

Zhu, Y., Qi, S., Zhang, B., He, D., Teng, Y., Hu, J., & Wei, X. (2019). Connectome-based biomarkers predict subclinical depression and identify abnormal brain connections with the lateral habenula and thalamus. *Frontiers in Psychiatry, 10,* 371.