Medical Image Analysis 90 (2023) 102932

Contents lists available at ScienceDirect

Medical Image Analysis

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

A-GCL: Adversarial graph contrastive learning for fMRI analysis to diagnose neurodevelopmental disorders☆

Shengjie Zhang [a,b], Xiang Chen [a,b], Xin Shen [c], Bohan Ren [d], Ziqi Yu [a,b], Haibo Yang [a,b], Xi Jiang [e], Dinggang Shen [f,g,h], Yuan Zhou [i,*], Xiao-Yong Zhang [a,b,**]

[a] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, 200433, China
[b] MOE Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, and MOE Frontiers Center for Brain Science, Fudan University, Shanghai, 200433, China
[c] Department of Mathematics, Beijing Normal University, Beijing, 100032, China
[d] Department of School of Cyber Science and Technology, Beihang University, Beijing, 100191, China
[e] Clinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for Neuroinformation, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, 611731, China
[f] School of Biomedical Engineering, ShanghaiTech University, Shanghai, 201210, China
[g] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, 200030, China
[h] Shanghai Clinical Research and Trial Center, Shanghai, 201210, China
[i] School of Data Science, Fudan University, Shanghai, 200433, China

ARTICLE INFO

Keywords:
Graph neural network
Adversarial graph contrastive learning
fMRI analysis
Neurodevelopmental disorders

ABSTRACT

Accurate diagnosis of neurodevelopmental disorders is a challenging task due to the time-consuming cognitive tests and potential human bias in clinics. To address this challenge, we propose a novel adversarial self-supervised graph neural network (GNN) based on graph contrastive learning, named A-GCL, for diagnosing neurodevelopmental disorders using functional magnetic resonance imaging (fMRI) data. Taking advantage of the success of GNNs in psychiatric disease diagnosis using fMRI, our proposed A-GCL model is expected to improve the performance of diagnosis and provide more robust results. A-GCL takes graphs constructed from the fMRI images as input and uses contrastive learning to extract features for classification. The graphs are constructed with 3 bands of the amplitude of low-frequency fluctuation (ALFF) as node features and Pearson's correlation coefficients (PCC) of the average fMRI time series in different brain regions as edge weights. The contrastive learning creates an edge-dropped graph from a trainable Bernoulli mask to extract features that are invariant to small variations of the graph. Experiment results on three datasets — Autism Brain Imaging Data Exchange (ABIDE) I, ABIDE II, and attention deficit hyperactivity disorder (ADHD) — with 3 atlases — AAL1, AAL3, Shen268 — demonstrate the superiority and generalizability of A-GCL compared to the other GNN-based models. Extensive ablation studies verify the robustness of the proposed approach to atlas selection and model variation. Explanatory results reveal key functional connections and brain regions associated with neurodevelopmental disorders.

1. Introduction

As the brain functional connectivity (FC) extracted from resting-state functional Magnetic Resonance Imaging (rs-fMRI) could reveal abnormal brain functional connections, it has been widely utilized in diagnosis of neurodevelopmental disorders, such as autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) (Canario et al., 2021). Currently, the diagnosis of these psychiatric diseases mainly relies on a subjective evaluation of abnormal behaviors by clinical experts (Hull et al., 2017). These cognitive and psychiatric assessments may contain intra- and inter-observer variability (Di and Biswal, 2020).

Functional magnetic resonance imaging (fMRI) provides a non-invasive way to observe cognitive and affective processes by measuring the FC between brain regions via blood oxygen level-dependent (BOLD) signals that dynamically reveal the change of brain functional connections (Chen et al., 2017; Chong et al., 2019). Since fMRI data serve as

☆ The first three authors contribute equally to this work.
* Corresponding author at: School of Data Science, Fudan University, Shanghai, 200433, China.
** Corresponding author at: Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, 200433, China.
E-mail addresses: yuanzhou@fudan.edu.cn (Y. Zhou), xiaoyong_zhang@fudan.edu.cn (X.-Y. Zhang).

https://doi.org/10.1016/j.media.2023.102932
Received 19 February 2023; Received in revised form 6 July 2023; Accepted 8 August 2023
Available online 22 August 2023

quantitative measures of brain functions, they may be used for accurate diagnosis while avoiding observer-related variability. Furthermore, by investigating the differences in fMRI between patients and normal controls (NC), they may reveal disease-specific changes within the brain (Canario et al., 2021).

1.1. Related work

Recently, computer-aided diagnosis (CAD) using rs-fMRI data has received increasing interest in the community (Bessadok et al., 2022). The main strategies include machine learning or deep learning-based methods, such as support vector machine (SVM) (Liu et al., 2020), random forest (RF) (Cordova et al., 2020), multilayer perceptron (MLP) (Hossain et al., 2021; Eslami and Saeed, 2019), convolutional neural networks (CNNs) (Heinsfeld et al., 2018), and convolution-based autoencoder (Almuqhim and Saeed, 2021). For example, Wang et al. conducted the classification by features derived from class-shared and class-specific decomposition (Wang et al., 2022a). Based on features extracted from fMRI data, several deep learning-based studies have been reported (Eslami et al., 2019; Heinsfeld et al., 2018; Yao et al., 2019; Kam et al., 2017; Parisot, 2018; Kazi et al., 2019; Chen et al., 2022; Jiang et al., 2020). Among them, graph neural networks (GNNs) have become an attractive framework for modeling brain networks due to their powerful graph embedding capabilities (Parisot, 2018; Kazi et al., 2019; Chen et al., 2022; Jiang et al., 2020).

In general, these GNN-based models can be divided into two categories: *graph classification* and *node classification*. Graph classification uses GNN as a nonlinear function that takes a graph from an fMRI image as input and outputs a class label. For example, BrainGB provides a standard pipeline including node feature construction, message passing, and graph pooling for brain network analysis (Cui et al., 2022a). BrainGNN (Li et al., 2021) leverages pooling regularization to extract the graph-level representation. A globally shared mask was leveraged to enhance the rs-fMRI classification capability of GNN backbones, achieving promising results among several datasets (Cui et al., 2022b). Pooling regularized GNN (Li et al., 2020) used an edge pooling strategy to remove some edges during the message passing process. Kim et al. proposed a spatio-temporal attention GNN (Kim et al., 2021), which combined GNN and Transformer to enhance the representation capability. Chen et al. conducted graph classification based on a node-edge graph attention network (NEGAT) (Chen et al., 2022). NEGAT combines structural magnetic resonance imaging (sMRI) data and rs-fMRI data to construct the graph. Inception-GCN (Kazi et al., 2019) utilized a receptive field-aware graph convolutional network to predict the disease.

On the other hand, node classification combines all the data to form a single population graph where each node is an instance that corresponds to an fMRI image and needs to be classified. For example, Cao et al. introduced non-imaging data – site, gender, IQ, age – into edge weight calculation (Cao et al., 2021), followed by 16 residual GNN layers for node classification. Hi-GCN (Jiang et al., 2020) leveraged a hierarchical graph convolutional network (GCN) to classify the graph representations embedded as node features. Zhou et al. proposed a graph-in-graph network (Zhou and Zhang, 2021), leveraging features extracted from a GCN along with non-imaging data to create the population graph. Zhang et al. constructed a local-to-global GNN for brain disorder classification using rs-fMRI, which increased the number of nodes in the population graph by applying different atlases (Zhang et al., 2022). Then, the nodes are classified by semi-supervised learning using a GCN layer.

With the advent of contrastive learning that achieved promising performance in natural image classification (He et al., 2020), *graph contrastive learning* (GCL) (Sun et al., 2021) has emerged as an effective approach to cope with rs-fMRI data. For each graph, by treating an augmented version of this graph as a *positive sample* and the other graphs as *negative samples*, GCL tries to learn representations that are close to the original graph in the embedding space for positive samples and far away from the original graph for negative samples. GCL has been applied in both graph and node classification tasks in rs-fMRI analysis. Yang et al. studied contrastive learning based pre-training of GNNs for brain network analysis and provided a few options of contrastive learning objectives (Yang et al., 2023). Peng et al. proposed graph canonical correlation analysis for temporal self-supervised learning (GATE) (Peng et al., 2022) that leveraged sliding windows to build a positive sample of an fMRI image, followed by semi-supervised learning to conduct node classification. Wang et al. proposed contrastive graph learning (CGL) (Wang et al., 2022b) that utilized contrastive learning to create node features, followed by node classification using a population graph. The proposed CGL calculated correlations of truncated BOLD signals to build positive samples. All the aforementioned methods leveraged the BOLD signals to build positive samples for contrastive learning. Another way is to take advantage of an additional GNN encoder for contrastive learning. Similar to the pooling regularized GNN (Li et al., 2020), a hierarchical signed graph pooling model was proposed to use two edge pooling strategies to generate features for contrastive learning (Tang et al., 2022).

Although these GNN models have achieved promising results in brain disease diagnosis using rs-fMRI, they still suffer from several problems. First, for node classification, some methods use additional non-imaging information to construct the population graph (Cao et al., 2021; Zhou and Zhang, 2021). However, when a new test instance with non-imaging features emerges, we need to calculate its similarity to all existing nodes in the graph in order to update the population graph. Consequently, the weights of the GNN must be updated to incorporate the new instance. This complexity in the testing process renders it more intricate compared to traditional graph classification algorithms. In addition, some instances may have missing non-imaging information (Bessadok et al., 2022), which hinders its incorporation into the population graph. Second, for graph classification, most of the current methods only use the information from the adjacency matrix, ignoring constructing reasonable node features, leading to unsatisfactory performance. Third, the creation of positive samples in current self-supervised models may not be ideal. For example, GATE (Peng et al., 2022) and CGL (Wang et al., 2022b) assumed that a graph generated from some part of the BOLD signals could be used as the positive sample. However, the arbitrary truncation of the BOLD signals might create a graph inferior to that obtained from the complete signals (Soon et al., 2021; Yan et al., 2020). To solve these problems, we propose a new GNN framework for robust rs-fMRI analysis to diagnose neurodevelopmental disorders.

1.2. Contribution

Inspired by graph contrastive learning (Suresh et al., 2021; You et al., 2020; Xu et al., 2021), we propose an adversarial graph contrastive learning (A-GCL) framework to conduct binary classification based solely on rs-fMRI data. To facilitate clinical applications, we use graph classification, which means that the rs-fMRI data of one patient leads to an individual graph. Instead of using traditional construction of node features (Cui et al., 2022a), three different frequency bands of the *amplitude of low-frequency fluctuation* (ALFF) (Yang et al., 2007) are calculated from the BOLD signals (Bu et al., 2019; Chen et al., 2023, 2022) and concatenated as the node features. These features are updated in a message-passing process. Then, the updated features are transformed to a mask that is employed to conduct adaptive edge dropping, with probability following a Bernoulli distribution. We call this mask a *Bernoulli mask*. The parameters in this process are trained by a contrastive policy with an adversarial loss (Kim et al., 2020). Compared with current state-of-the-art (SOTA) methods, A-GCL uses the original graph and its edge-dropped version for contrastive learning and feature distillation. The distilled features can be classified by a linear classifier for disease diagnosis. The remaining edge weights

after the trained Bernoulli mask can be analyzed as biomarkers for the disease, similar to the interpretation strategy in mask-guided GNN which does not use contrastive learning (Cui et al., 2022b).

In contrast to truncating the BOLD signals, we use the complete BOLD signals to construct a graph and then learn a Bernoulli mask from this graph to randomly drop edges to generate a positive sample in contrastive learning. The idea of this Bernoulli mask is motivated by the success of masked autoencoder (MAE) (He et al., 2022) in vision tasks. In MAE, some patches of the input image are randomly masked out, while the remaining unmasked patches are utilized to represent the entire image. This strategy is based on the assumption that redundant information exists in the original image, and self-supervised learning can effectively eliminate such information. Similarly, in our work, we adopt a similar perspective by assuming the presence of redundant information within the adjacency matrix of the graph obtained from the fMRI data. Thus, we aim to leverage contrastive learning to remove this redundant information and enhance the representation of the graph.

Another notable aspect of our approach is the implementation of a *dynamic memory bank*, which enables the collection of diverse negative sample features. Typically, a memory bank contains negative sample features from the same batch. However, the performance of contrastive learning is highly dependent on the number of negative samples (He et al., 2020). To increase the number of negative samples, one approach is to combine samples from all the training batches. Nevertheless, such a practice would lead to too much GPU memory usage. In this work, we use a queue to store sample features from the same batch as well as sample features from different batches. The sample features within the queue are dynamically updated during iterations. This enables us to augment the diversity of negative samples while effectively managing the GPU memory overhead.

We validated our proposed method on the Autism Brain Imaging Data Exchange (ABIDE) I, ABIDE II, and ADHD-200 datasets. Extensive experiments on three different atlases – AAL1, AAL3, Shen268 – show that the A-GCL framework outperforms other SOTA methods significantly. The major contributions of our work are summarized as follows:

1 **Adversarial contrastive learning with a dynamic memory bank:** A-GCL performs contrastive learning on features extracted from the original graph and its edge-dropped version based on a Bernoulli mask. This enables the learned features to be independent of the class labels and to genuinely represent an embedding of the graphs in the Euclidean space. Furthermore, a dynamic memory bank is implemented to further enhance this feature extraction process.
2 **Multiple datasets and atlases:** A-GCL is evaluated on 2 neurodevelopmental diseases, 3 rs-fMRI datasets, with 3 different atlases. A-GCL achieved the best performance when compared with other competing models, demonstrating its superiority and generalizability.
3 **Ablation study:** Extensive ablation studies including transfer learning of ABIDE I to ABIDE II, the influence of the embedding dimension, the GNN encoder, the graph augmentation strategy, the adversarial training strategy, the edge weights, and node features, are conducted to verify the robustness of A-GCL.
4 **Explanation:** Explanatory analyses including visualization of the Bernoulli mask and important ROIs are performed to identify brain regions associated to the diseases. These brain regions could be used as biomarkers in further understanding these diseases (ASD and ADHD).

2. Method

The framework of our proposed A-GCL is shown in Fig. 1. First, two kinds of information – an adjacency FC matrix representing edge connections and a set of ALFF node features – are extracted from the fMRI data to form a graph. The graph is then fed into the A-GCL network to produce latent features that will be used for classification. Finally, the related brain FCs and important regions are analyzed for interpretation.

2.1. Graph construction

The graph is constructed as shown in Fig. 1(a). Given an atlas, the fMRI images are parcellated into many ROIs. Each ROI is considered as a node and the functional connectivity between any pair of these ROIs is considered as edges to form a graph. In each ROI, the mean time series is calculated by averaging all the BOLD signals in the region. The edge weights are calculated by Pearson's correlation coefficient between the mean time series of two regions. The node features are derived from 3 frequency bands of the ALFFs (Slow-5: 0.01–0.027 Hz, Slow-4: 0.027–0.073 Hz, classical: 0.01–0.08 Hz) in BOLD signals, which are defined as the total power within the low-frequency range and are calculated from the Fourier transform of the mean time series (Guo et al., 2017). Note that the effectiveness of ALFFs as node features has been demonstrated in several previous studies (Bu et al., 2019; Chen et al., 2023, 2022).

The resulting graph is denoted by $G = (V, A, X, E)$, where $V = \{v : v = 1, \dots, M\}$ represents the set of nodes, $A = [a_{uw} : u, w \in V] \in \{0,1\}^{M\times M}$ represents the adjacency matrix indicating if there exists an edge between two nodes ($a_{uw} = 1$), $X = \{x_v \in \mathbb{R}^3 : v \in V\}$ represents the set of node features and $E = [e_{uw} \in \mathbb{R} : u, w \in V] \in \mathbb{R}^{M\times M}$ the matrix of edge weights, M is the number of nodes/ROIs. The adjacency matrix is initialized with all 1's. The node features are normalized to $[0, 1]$ by subtracting the minimum from all the 3 channels and dividing the result by the difference between the maximum and the minimum. The edge weights are normalized to $[-1, 1]$ by dividing each weight by the maximum of the absolute values.

2.2. A-GCL

Our proposed A-GCL includes graph augmentation, random edge dropping using a Bernoulli mask, weight-shared GNN encoders, and the loss function.

2.2.1. Graph augmentation

Let $G = (V, A, X, E)$ represent the graph constructed from the fMRI data. The graph goes through a *graph isomorphism network* (GIN) block, a feature concatenation, and an MLP layer to form an augmented graph for learning features that are invariant to small variations.

The GIN block consists of 2 layers. Each layer updates the node features by using a message-passing process. Let the set of neighboring nodes of a certain $v \in V$ be denoted by $\mathcal{N}_v$, the message passing process proceeds according to

$$h_v^{(k)} = g^{(k)}(h_v^{(k-1)}, f^{(k)}(\{(h_u^{(k-1)}, e_{uv}) : u \in \mathcal{N}_v\})),\ k = 1, 2$$

where $h_v^{(0)} = x_v$, $f^{(k)}$ is a function that transforms the neighboring node features and edge weights to an aggregated vector. $g^{(k)}$ is a trainable function that maps the current node representation and the aggregated vector to a new representation. Here, $f^{(k)}$ is the weighted sum of node features and edge weights, $g^{(k)}$ is a MLP layer. Thus the message-passing process is:

$$h_v^{(k)} = MLP^{(k)}(h_v^{(k-1)} + \sum_{u\in\mathcal{N}_v} h_u^{(k-1)} e_{uv}).$$

The above process can be written in a matrix form:

$$H^{(k-\frac{1}{2})} = (I + A \circ E)H^{(k-1)},$$

$$H^{(k)} = \mathrm{BN}(\sigma_{ReLU}(H^{(k-\frac{1}{2})}W_1^{(k)} + \mathbf{1}b_1^{(k)})W_2^{(k)} + \mathbf{1}b_2^{(k)}),$$

where $H^{(k)}$ is a matrix with rows being $\{h_v^{(k)}\}$, $\circ$ denotes the Hadamard product (element-wise multiplication), $\mathbf{1}$ is a M-dimensional vector

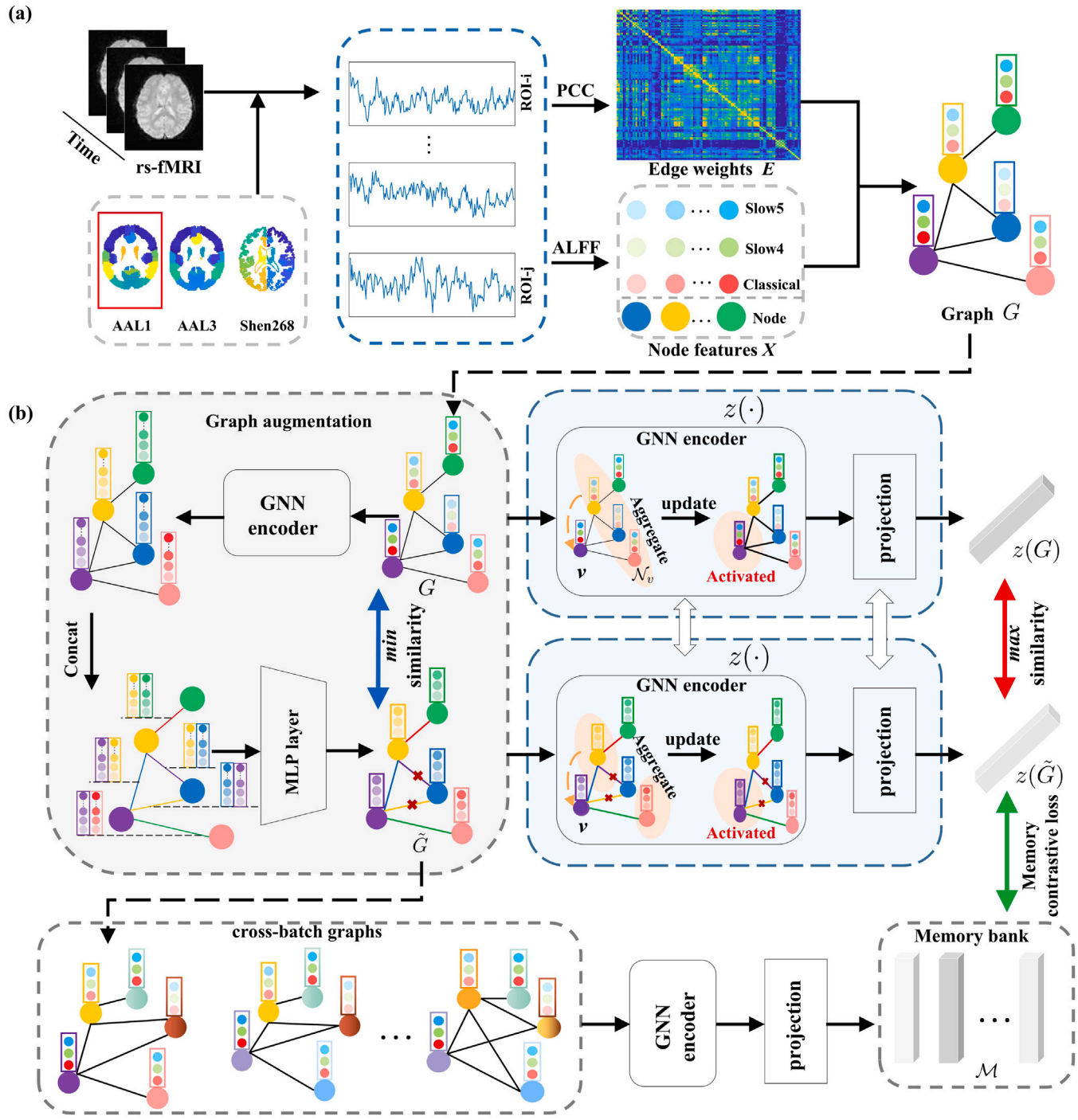


Fig. 1. The overview of our proposed A-GCL. (a) shows the construction of node features and edge weights, including atlas selection, ALFF calculation, Pearson's correlation coefficient (PCC) calculation, and the structure of the graph. (b) shows the framework of A-GCL, including random graph augmentation from a trainable Bernoulli mask, a weight-shared GNN encoder, and a projection head. The purpose of A-GCL is to extract latent feature vectors using contrastive learning.

consisting of all 1's, $W_1^{(k)} \in \mathbb{R}^{d^{(k-1)}\times d}$, $b_1^{(k)} \in \mathbb{R}^{1\times d}$ and $W_2^{(k)} \in \mathbb{R}^{d\times d}$, $b_2^{(k)} \in \mathbb{R}^{1\times d}$ are trainable parameters with $d^{(0)} = 3$ and $d^{(1)} = d$, BN denotes a batch normalization operation. The GIN block aims to learn a latent representation of node features $\tilde{x}_v = h_v^{(2)}$, which are concatenated to create embedded edge features $\tilde{x}_{uv} = [\tilde{x}_u; \tilde{x}_v] \in \mathbb{R}^{2d}$.

The edge features are fed into an MLP layer to generate parameters of a Bernoulli distribution that is sampled to randomly drop edges. The MLP layer consists of a linear layer with trainable weights $W_3 \in \mathbb{R}^{2d\times 2d}$, $b_3 \in \mathbb{R}^{1\times 2d}$, a ReLU activation function σ_{ReLU}, another linear layer with weights $W_4 \in \mathbb{R}^{2d\times 1}$, $b_4 \in \mathbb{R}$ and finally a sigmoid function that converts the number to the range $(0, 1)$:

$$\mu_{uv} = \sigma_{Sigmoid}(\sigma_{ReLU}(\tilde{x}_{uv}^{\top} W_3 + b_3)W_4 + b_4).$$

In this way, the edge features are transformed into a scalar that corresponds to the parameters of a Bernoulli distribution.

Given the set of parameters $\{\mu_{uv}\}$, an indicator for edge dropping b_{uv} is sampled for each edge, i.e. $b_{uv} \sim Bernoulli(\mu_{uv})$, where $b_{uv} = 0$ indicates that the edge will be dropped. The matrix $[b_{uv} : u, v \in V]$ constitutes the binary (0 or 1) *Bernoulli mask B*. To enable the gradient to backpropagate, the sampling process needs to be reparameterized.

We use the following reparameterization trick (Luo et al., 2020):

$$b_{uv} = \sigma_{Sigmoid}((\log \frac{\epsilon_{uv}}{1-\epsilon_{uv}} + \log \frac{\mu_{uv}}{1-\mu_{uv}})/\tau),$$

where $\epsilon_{uv} \sim Uniform(0,1)$, and τ denotes a temperature parameter that controls the smoothness of the reparameterized sampling functions. The term inside the sigmoid function is positive if ϵ_{uv} is in the range $(1-\mu_{uv}, 1)$ and negative if it is in the range $(0, 1-\mu_{uv})$. Hence, the output b_{uv} approaches 1 with probability μ_{uv}. As $\tau \to 0$, b_{uv} gets closer to the sampled binary indicator.

The Bernoulli mask is applied to the adjacency matrix A by element-wise multiplication, hence masks out some edges with $b_{uv} = 0$. Thereby, the input graph becomes an edge-dropped graph with the same number of nodes. Denoting the process that creates the matrix of parameters $[\mu_{uv}]$ from graph $G = (V, A, X, E)$ as μ, the whole data augmentation process can be written as

$$B \sim Bernoulli(\mu(G)), \quad \tilde{G} = (V, A \circ B, X, E). \tag{1}$$

2.2.2. Dynamic memory bank and loss function design

The input graph G and the augmented graph $\tilde{G}$ derived above are input into a graph encoder and a projection head to extract features. As shown in Fig. 1(b), they go through a weight-shared GIN block and a weight-shared projection head. The GIN block has the same architecture as the one in the graph augmentation. After the GIN block, the graph is converted to a vector by summing up all the node features. This vector goes through a projection head implemented by a 2-layer MLP. The MLP has 2 dimension-preserving linear layers with a ReLU activation function in the middle. Denote the GIN block and the projection head combined feature extractor by z, we can get two feature vectors $z(G), z(\tilde{G}) \in \mathbb{R}^d$ after the projection head.

To train the feature extractor and the parameters in the graph augmentation process, we use a loss function that forces the two feature vectors to be close if they are from the same graph and far away if they are from different graphs. In this way, the trained feature extractor can keep the most important information while removing excessive information in the graph. Since the training proceeds in a batch-wise manner, such a loss is also imposed batch-wisely. Specifically, the loss is defined by the infoMax principle (Veličković et al., 2018) which we aim to maximize:

$$I(z, \mu; \mathcal{B}) = \frac{1}{|\mathcal{B}|} \sum_{G \in \mathcal{B}} \log \frac{\exp(\mathrm{sim}(z(G), z(\tilde{G})))}{\sum_{G' \in \mathcal{B} \setminus \{G\}} \exp(\mathrm{sim}(z(G), z(\tilde{G}')))}$$

where the dependence of $\tilde{G}$ (resp. $\tilde{G}'$) on G (resp. G') is given in Eq. (1), $\mathcal{B}$ is the set of graphs in a batch, $|\mathcal{B}|$ is its cardinality, sim is the similarity metric and we use the cosine of the angle between two input vectors:

$$\mathrm{sim}(z_1, z_2) = z_1^\top z_2 / (\|z_1\| \|z_2\|).$$

Maximizing $I(z, \mu; \mathcal{B})$ could be easily achieved by retaining all the edges. To force more edges to be dropped in this context, we need a regularization term to facilitate edge dropping. Such a regularization term is designed to be the mean of all the μ_{uv}'s

$$R(\mu; \mathcal{B}) = \frac{1}{|\mathcal{B}| M^2} \sum_{G \in \mathcal{B}} \mathbf{1}^\top \mu(G) \mathbf{1},$$

and we want to minimize $R(f; \mathcal{B})$.

To further encourage extracting features that vary smoothly on the graph manifold, we build a cross-batch memory bank $\mathcal{M}$ that stores features $z(\tilde{G}')$ from previous batches. $\mathcal{M}$ has a preset length and follows the first-in-first-out (FIFO) rule. An additional loss function using the memory bank is maximized:

$$I(z, \mu; \mathcal{B}, \mathcal{M}) = \frac{1}{|\mathcal{B}|} \sum_{G \in \mathcal{B}} \log \frac{\exp(\mathrm{sim}(z(G), z(\tilde{G})))}{\sum_{z' \in \mathcal{M}} \exp(\mathrm{sim}(z(G), z'))}.$$

where the dependence of $\tilde{G}$ on G is given in Eq. (1). The memory bank $\mathcal{M}$ is initialized with zeros and updated with the iteration of the batches.

Table 1
Demographic information of studied subjects in ABIDE I, ABIDE II, and ADHD. ASD: autism spectrum disorder, NC: normal control, ADHD: attention deficit hyperactivity disorder.

Dataset	ABIDE I		ABIDE II		ADHD-200	
Information	ASD	NC	ASD	NC	ADHD	NC
Subject	467	520	243	289	215	291
Gender (F/M)	63/404	97/423	40/203	105/184	59/156	158/139
Age (mean ± std)	16.5 ±8.3	16.8 ±7.7	13.6 ±8.4	12.6 ±7.4	11.4 ±3.0	12.0 ±3.3

Finally, with regularization coefficients λ_1 and λ_2, the objective function is defined as

$$\min_{\mu} \max_{z} I(z, \mu; \mathcal{B}) + \lambda_1 R(\mu; \mathcal{B}) + \lambda_2 I(z, \mu; \mathcal{B}, \mathcal{M}). \tag{2}$$

The optimal μ and z are obtained through gradient descent/ascent of the corresponding parameters.

2.3. Classification and interpretation

The representation ability of the trained latent feature vectors $z(G)$ is validated in the diagnosis of ASD or ADHD. Here, a simple linear classifier, SVM, is used to classify the extracted features. After training the classifier, we also try to discover brain ROIs closely associated with the diseases.

Since the extracted latent features are important for interpretation, we interpret the proposed A-GCL by the trained Bernoulli mask. This mask leads to an augmented sparse graph which can be interpreted in two ways. First, the important connections in this sparse graph can be visualized. Second, the importance score of a node can be calculated by summing up the elements in a row of $A \circ B \circ E$, as the larger the summation, the more connections remain after dropping edges. This importance score can be seen as the degree of connection to diseases of the corresponding brain region.

2.4. Implementation details

Our model is implemented in PyTorch. All of the algorithms described in this paper can be executed on a single GPU. The experiments are accelerated by two servers with 8 NVIDIA V-100 GPUs and 2 NVIDIA A-6000 GPUs. Our code – including pre-processing (Matlab), A-GCL training, and evaluation scripts (Python 3.7) – has been released at https://github.com/qbmizsj/A-GCL. The implementation details are as follows: The learning rate is set to 0.0005. The embedded dimension d is set to 32. The batch size is 32. The temperature τ is set to 1. The regularization coefficient λ_1 and λ_2 are set to 2 and 0.4, respectively. The length of the memory bank $\mathcal{M}$ is set to 256. When applying the proposed A-GCL to a new fMRI dataset, it is recommended to search the batch size in {8, 16, 32, 64}, the learning rate of the optimization over the parameters of μ in {0.0001, 0.0005, 0.001, 0.005, 0.01}, and the learning rate of the optimization over the parameters of z in {0.0005, 0.001, 0.01}.

3. Results

3.1. Experimental setup

3.1.1. Dataset and preprocessing

We use three rs-fMRI datasets – Autism Brain Imaging Data Exchange (ABIDE) I, ABIDE II, and ADHD-200 – which are publicly available MRI datasets collected from different international imaging sites. The ground truth labels can be accessed from the phenotypic file when downloading the datasets. As described in the dataset documentation, these labels are derived through meticulous diagnostic procedures and

Table 2
Results for ASD classification (ASD vs. NC) on ABIDE I and ABIDE II, and those for ADHD classification (ADHD vs. NC) on ADHD-200 using the AAL1 atlas. The results (in %) were calculated based on 5-fold cross-validation. The training time (trn time) for each epoch (in 's') and the inference time (inf. time) for each sample (in 'ms') are also included. The best result in each category is highlighted in **red**.

Dataset	Method	Trn time	Inf. time	Accuracy	AUC	Precision	Recall	F1-score	Avg
ABIDE I	MLP	7.59	0.75	63.20 ± 4.62	64.03 ± 3.89	64.58 ± 3.25	65.73 ± 2.80	65.15 ± 3.20	64.54 ± 3.55
	SVM	2.14	0.04	66.37 ± 3.82	64.08 ± 3.19	62.30 ± 4.83	70.57 ± 2.84	66.18 ± 3.45	65.90 ± 3.63
	v-GCN	0.99	0.34	69.40 ± 3.54	70.79 ± 3.12	68.50 ± 2.88	76.07 ± 3.43	72.10 ± 3.09	71.37 ± 3.21
	GraphSage	1.04	0.37	71.13 ± 3.45	70.41 ± 3.23	72.44 ± 2.74	73.05 ± 2.91	72.75 ± 2.86	71.96 ± 3.04
	GIN	1.06	0.40	70.08 ± 3.69	70.50 ± 3.41	71.19 ± 3.05	73.82 ± 2.90	72.48 ± 2.94	71.61 ± 3.20
	HI-GCN	4.58	0.87	70.59 ± 3.36	71.42 ± 2.90	68.47 ± 3.47	73.01 ± 2.85	70.67 ± 3.13	70.83 ± 3.14
	AL-NEGAT	6.20	0.52	71.04 ± 3.50	72.40 ± 2.69	75.59 ± 2.80	70.27 ± 3.34	72.83 ± 3.05	72.43 ± 3.08
	BrainGNN	9.21	0.27	68.29 ± 3.87	70.72 ± 3.16	66.15 ± 3.14	71.78 ± 3.20	68.85 ± 3.16	69.19 ± 3.31
	DGCN	12.79	0.42	73.30 ± 3.02	74.15 ± 3.08	72.06 ± 2.68	73.55 ± 3.19	72.81 ± 2.84	73.17 ± 2.96
	GATE	4.55	0.18	73.52 ± 3.16	75.60 ± 2.84	74.36 ± 3.04	75.60 ± 2.84	74.60 ± 3.06	74.74 ± 2.99
	A-GCL	2.14	0.20	80.65 ± 2.88	81.42 ± 2.85	80.02 ± 2.94	82.28 ± 3.10	81.14 ± 2.96	81.10 ± 2.95
ABIDE II	MLP	7.62	0.74	61.18 ± 5.16	62.85 ± 5.15	62.87 ± 4.12	64.35 ± 3.50	63.60 ± 3.85	62.97 ± 4.36
	SVM	2.16	0.04	64.48 ± 4.29	62.81 ± 4.80	60.56 ± 7.28	70.24 ± 5.40	65.04 ± 6.26	64.63 ± 5.61
	v-GCN	0.99	0.32	70.05 ± 4.22	71.48 ± 5.26	70.13 ± 3.74	74.28 ± 3.90	72.14 ± 3.87	71.62 ± 4.20
	GraphSage	1.06	0.39	71.17 ± 3.85	70.86 ± 3.64	72.35 ± 3.45	73.00 ± 3.40	72.67 ± 3.44	72.01 ± 3.56
	GIN	1.08	0.40	68.56 ± 3.90	70.20 ± 4.06	70.46 ± 4.09	72.16 ± 3.84	71.30 ± 3.92	70.54 ± 3.96
	HI-GCN	4.65	0.90	70.62 ± 4.19	70.86 ± 3.57	69.10 ± 4.33	71.88 ± 3.17	70.46 ± 3.62	70.58 ± 3.78
	AL-NEGAT	6.14	0.58	70.02 ± 4.64	71.40 ± 3.19	73.62 ± 4.14	71.25 ± 5.05	72.42 ± 4.30	71.74 ± 4.26
	BrainGNN	8.78	0.26	66.25 ± 6.72	67.48 ± 5.05	69.59 ± 4.61	66.50 ± 3.92	68.01 ± 4.25	67.57 ± 4.91
	DGCN	12.46	0.48	72.58 ± 3.38	73.27 ± 3.41	72.31 ± 3.30	73.25 ± 4.61	72.78 ± 3.64	72.84 ± 3.67
	GATE	4.50	0.20	72.09 ± 4.50	74.06 ± 4.73	74.18 ± 3.77	73.09 ± 3.45	73.63 ± 3.86	73.41 ± 4.06
	A-GCL	2.15	0.18	79.88 ± 3.48	80.04 ± 3.37	79.14 ± 3.24	81.47 ± 4.60	80.29 ± 3.64	80.16 ± 3.67
ADHD	MLP	8.03	0.87	61.08 ± 5.61	62.83 ± 3.50	62.31 ± 5.37	63.74 ± 4.27	63.01 ± 4.62	62.59 ± 4.67
	SVM	3.65	0.06	62.20 ± 4.65	60.74 ± 5.32	60.79 ± 5.30	65.55 ± 4.83	63.08 ± 5.18	62.47 ± 5.06
	v-GCN	1.21	0.44	62.85 ± 3.48	64.40 ± 3.41	63.06 ± 4.40	64.73 ± 5.09	63.88 ± 4.85	63.78 ± 4.25
	GraphSage	0.88	0.46	62.10 ± 4.42	62.29 ± 4.49	63.68 ± 5.10	65.76 ± 4.72	64.70 ± 4.83	63.71 ± 4.71
	GIN	0.81	0.35	60.35 ± 5.96	62.15 ± 3.76	62.48 ± 3.54	62.04 ± 4.13	62.36 ± 3.77	61.88 ± 4.23
	HI-GCN	5.12	0.57	63.30 ± 5.35	63.84 ± 5.17	65.88 ± 4.17	62.37 ± 5.15	64.08 ± 4.50	63.89 ± 4.87
	AL-NEGAT	6.68	0.25	64.25 ± 4.24	62.52 ± 4.10	63.55 ± 4.80	66.79 ± 4.36	65.13 ± 4.59	64.45 ± 4.42
	BrainGNN	9.04	0.48	62.75 ± 3.69	62.28 ± 5.83	65.08 ± 3.36	60.24 ± 4.12	62.57 ± 3.68	62.58 ± 4.14
	DGCN	10.85	0.57	63.38 ± 4.10	65.04 ± 5.08	64.25 ± 4.52	63.86 ± 5.24	64.06 ± 4.70	64.12 ± 4.73
	GATE	5.61	0.26	65.26 ± 3.75	65.72 ± 3.86	67.90 ± 5.72	66.08 ± 4.14	66.98 ± 4.61	66.39 ± 4.42
	A-GCL	1.32	0.18	70.92 ± 4.28	71.12 ± 4.45	72.57 ± 4.65	73.02 ± 4.03	72.79 ± 4.51	72.08 ± 4.38

clinical assessments, specifically utilizing the Autism Diagnostic Observation Schedule (ADOS) for ABIDE and the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for ADHD-200. The ABIDE I dataset consists of 1112 subjects: 539 ASD patients and 573 normal controls (NC). A reliable pipeline, *fMRIPrep* (Esteban et al., 2019), was used for preprocessing the fMRI images. Specifically, rs-fMRI reference image estimation, head-motion correction, slice timing correction, and susceptibility distortion correction are performed. For confounder removal, framewise displacement, global signals, and mean tissue signals are taken as the covariates and regressed out after registering the fMRI volumes to the standard MNI152 space. In this work, 467 ASD and 520 NC are included after quality checking based on DVARS (Power et al., 2012) and framewise displacement (Power, 2017). The ABIDE II (521 ASD and 593 NC samples) and ADHD-200 (362 ADHD and 585 NC) datasets also underwent the same pre-processing procedure and quality check. The number of subjects included in this study and their demographics are given in Table 1.

Three brain atlases – AAL1, AAL3 (Rolls et al., 2020), and Shen268 (Shen et al., 2013) – were used to parcellate the brain into 116, 166, and 268 regions, respectively. For each atlas, the mean time series (BOLD signal) in each region is calculated by averaging the time series of all the voxels. Then, the ALFF node features are calculated from the mean time series and the full FC matrix is calculated using Pearson's correlation coefficient (PCC) between two mean time series. Based on the 3 atlases, each subject produces 3 different graphs. We use the graphs from AAL1 for validating the classification performance and the graphs from AAL3 and Shen268 for validating the robustness of our framework to atlas selection.

3.1.2. Competing methods

We compare the proposed A-GCL with 10 machine learning methods as follows:

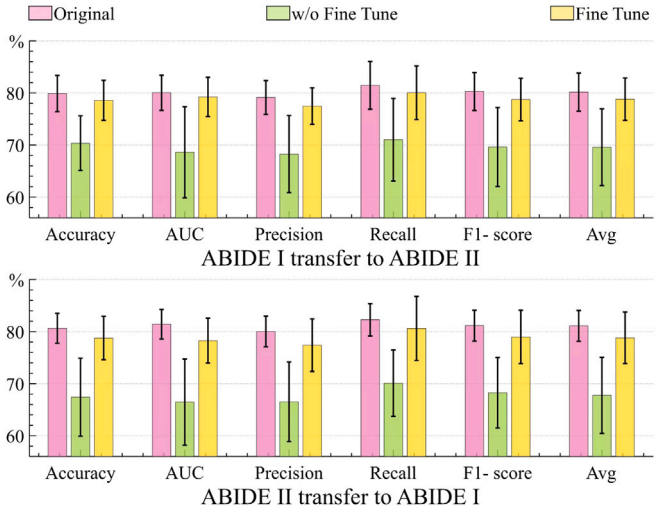


Fig. 2. Transfer learning of A-GCL on two ABIDE datasets. The result is shown in mean±std over 5-fold cross validation.

1. Traditional machine learning methods: FC and ALFFs are flattened and concatenated into a vector, which is fed to an MLP or SVM. To train the competing MLP, we set the learning rate to 0.01 and experimented with different numbers of nodes in the hidden layers, including {1000, 100}, {1000, 500}, and {1000, 500, 100}. The best numerical result of MLP was reported after evaluating these configurations. We use a linear SVM and set the penalty coefficient from {0.1,0.5,1,5,10}.

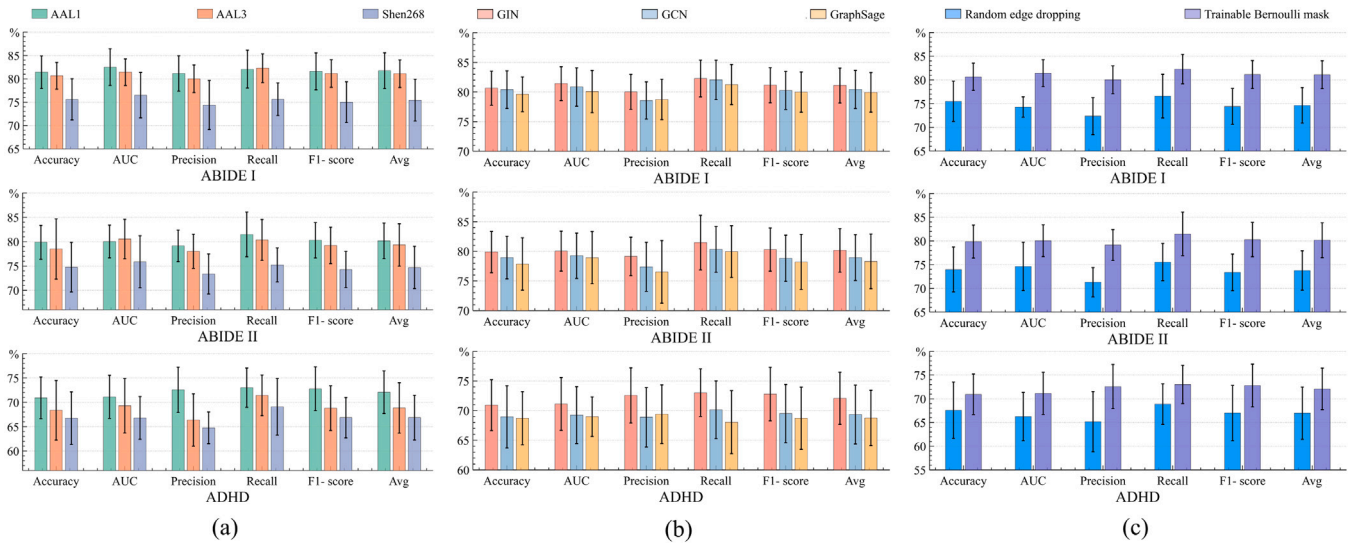


Fig. 3. An ablation study was conducted to investigate the impact of several factors on fMRI classification, including (a) atlases, (b) GNN encoders, and (c) edge-dropping strategies. The results are presented as mean ± std over a 5-fold cross validation.

2. Standard GNNs: vanilla GCN (v-GCN) (Kipf and Welling, 2016), GraphSage (Hamilton et al., 2017), graph isomorphism network (GIN) (Xu et al., 2018). We have included all the code implementation of the aforementioned GNN baselines in our released code. For the GNN models, we experiment with different learning rates including {0.0001, 0.0005, 0.001, 0.01} and vary the batch size from {8, 16, 32, 64}. We report the best results for each competing method.
3. Code-released works for fMRI analysis:
   - HI-GCN (Jiang et al., 2020) (https://github.com/haojiang1/hi-GCN). Adam is set as the optimizer with 0.001 as the learning rate. We search the learning rate from {0.0001, 0.0005, 0.001, 0.01}. The batch size is chosen from {8,16,32,64}. All the other hyper-parameters follow the original implementation of HI-GCN.
   - AL-NEGAT (Chen et al., 2022) (https://github.com/XiJiangLabUESTC). We replace the T1 intensity of AL-HEGAT with the ALFF node features that we use. We tried the number of GNN layers from 2 to 4, ϵ in AL-NEGAT from {0.001, 0.005, 0.01, 0.02}. All the remaining hyper-parameters are kept unchanged.
   - BrainGNN (Li et al., 2021) (https://github.com/xxlya/BrainGNN_Pytorch). We have tried two implementations of BrainGNN, the original implementation and the one with the node features replaced by our calculated ALFFs. We observe that the original BrainGNN failed to obtain satisfactory performance on the ABIDE datasets. Thus we report the results of the version with the node features replaced. We set all the remaining hyper-parameters as the original implementation.
   - DGCN (Zhao et al., 2022) (https://github.com/zhangyubrain/DGCN). We search the learning rate in {0.001, 0.005, 0.01} as the original implementation suggests a learning rate of 0.01 on ADHD. The number of graph convolutional layers varies from 2 to 6. The remaining hyper-parameters are set as the original implementation.
   - GATE (Peng et al., 2022) (https://github.com/LarryUESTC/GATE). As GATE provides the optimal combination of several hyper-parameters on the ABIDE dataset, we also take it as a reference on ADHD-200 and search the window size in {20, 30, 50, 70}, and the step size in {10,15,20,25,30}

3.1.3. Evaluation strategy

The performance of the A-GCL framework is evaluated by five metrics: accuracy, sensitivity, specificity, F1-score, and AUC. An average of these 5 metrics is also reported. To avoid the bias induced by a single split of the dataset, 5-fold cross-validation was employed.

3.2. Classification performance

The classification results on all the datasets are shown in Table 2. We can see that our proposed A-GCL achieves the highest mean accuracy (80.65%, 79.88%, 70.92%) on the three datasets by using the AAL1 atlas, which is about 5%–7% higher than the other SOTA methods. Two machine learning methods achieve the worst performance and GNN-based methods share similar performance in ASD or ADHD classification. It is noted that although GATE uses basic contrastive learning, it still achieves the second best performance among all the methods, demonstrating that contrastive learning is beneficial to the performance of fMRI classification. Another highlight is that our comparison results are collected from 5-fold cross validation in contrast to the previous works (Cao et al., 2021; Chen et al., 2022; Eslami et al., 2019; Heinsfeld et al., 2018; Yao et al., 2019; Parisot, 2018; Kazi et al., 2019; Kam et al., 2017) that only report the accuracy of a single split of the dataset.

3.2.1. Transfer learning for ABIDE datasets

To further investigate the generalizability of A-GCL, we perform transfer learning on the ABIDE datasets: train the model on ABIDE I and use the trained model directly (or after fine tuning) on ABIDE II, and vice versa. As shown in Fig. 2, a simple application using A-GCL trained on ABIDE I is not effective enough for ABIDE II. The accuracy declines by about 10% when compared with the accuracy derived from A-GCL trained from scratch on ABIDE II. However, fine tuning based on the previously trained parameters yields comparable performance to the GNN-based models in Table 2. In addition, when the model is trained on ABIDE II and applied to ABIDE I, a similar phenomenon is seen.

3.3. Ablation studies

3.3.1. Influence of different atlases on the three datasets

To investigate the robustness of A-GCL to atlas selection, we use three different atlases – AAL1, AAL3, and Shen268 – to construct graphs with different numbers of nodes. The classification results are reported in Fig. 3(a).

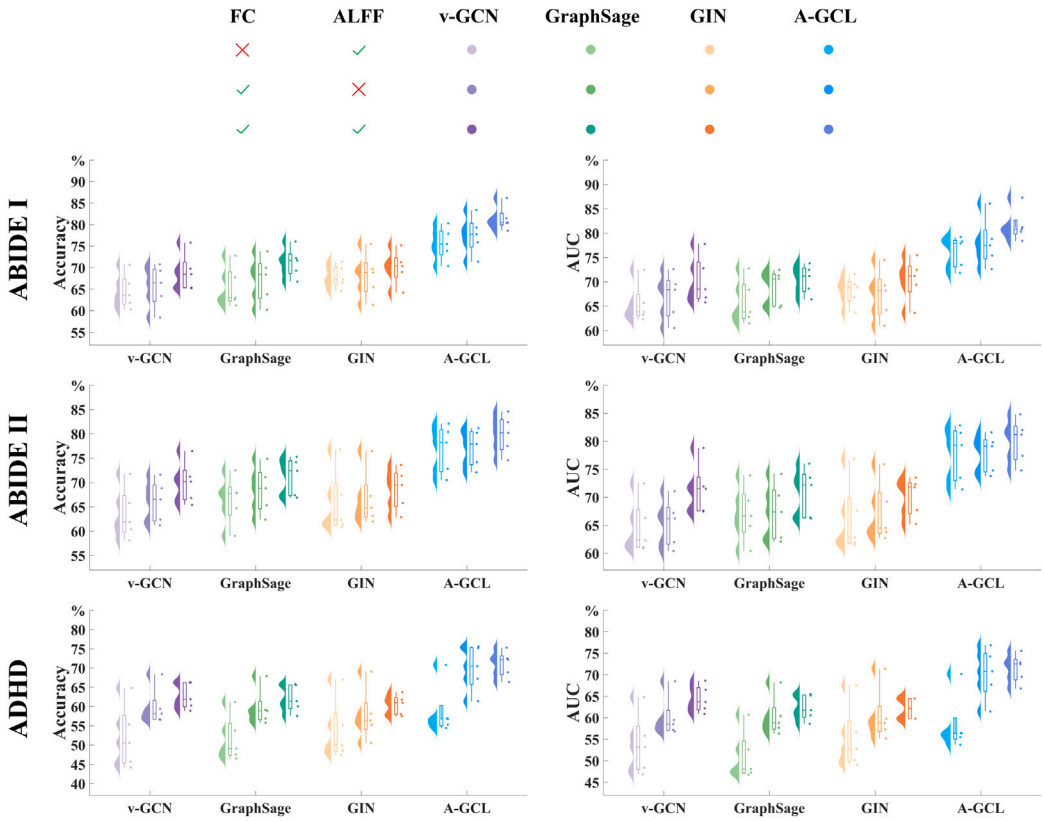


Fig. 4. Ablation study on how ALFF and FC influence the fMRI classification. The result is shown in mean±std over 5-fold cross validation.

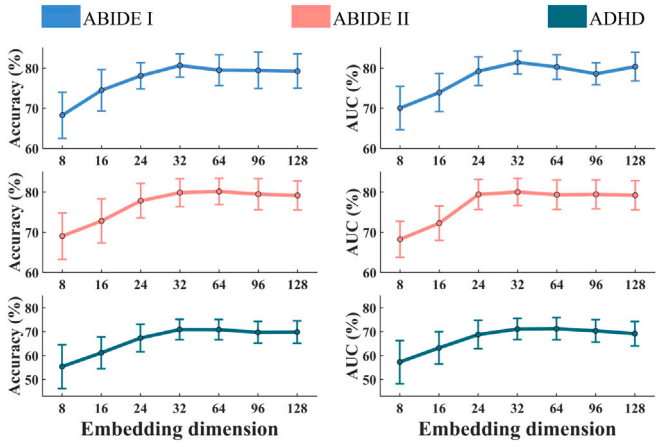


Fig. 5. Ablation study on whether the embedding dimension will influence the classification performance. The vertical lines denote the standard deviation. The result is calculated from 5-fold cross validation.

As we previously stated, we aim to reduce the redundancy of information between patients and the NC group through A-GCL. The performance achieved using the Shen268 atlas is not as good as that of the two AAL atlases, suggesting that more nodes in the graph can be detrimental to the performance. However, in the ASD classification task, the performance improves slightly on ABIDE I but deteriorates slightly on ABIDE II as the number of nodes increases from 116 (AAL1) to 166 (AAL3). In general, the performance achieved by the AAL atlases surpasses that of the Shen268 atlas across the three datasets, suggesting that overly elaborate brain atlases may not be ideal in fMRI classification tasks.

3.3.2. *Effectiveness of edge weights and node features*

We investigate the effectiveness of the edge weights and node features (ALFF) on the AAL1 atlas using four GNN-based models. To evaluate the effectiveness of the edge weights derived from the PCC calculation, we replace all elements of the FC matrix with 1, thus creating a version without edge weights. As shown in Fig. 4, compared to the version without edge weights, the version with edge weights improves the classification performance in all the GNNs, demonstrating that the edge weights do increase the accuracy.

As for the importance of node features, we replace the ALFF features with a vector of all 1's and consider this as a version without node features. The accuracy of the GNN-based methods without node features is also shown in Fig. 4. We can see that with the ALFF features included, the performance increases by 2% to 3% in all the methods on ABIDE I and ABIDE II. It is observed that the methods with only ALFF and without the FC matrix could achieve moderate results in ASD classification, but lead to much worse results when it comes to ADHD classification. Nevertheless, it is obvious that both ALFF and FC contribute to the highest score of accuracy and AUC. This phenomenon could be due to that without ALFF, GNN-based methods are prone to produce a worse representation during the message passing process. Hence, the ALFF node features and PCC edge weights indeed benefit the proposed A-GCL in improving the classification performance.

3.3.3. *Influence of the GNN encoder*

To study the influence of different GNN encoders on the classification performance of A-GCL, we replace the GIN block with GCN and GraphSage and evaluate the performance on the three datasets using the AAL1 atlas. As shown in Fig. 3(b), three GNN encoders share similar performance in terms of accuracy, specificity, sensitivity, and F1-score, suggesting that A-GCL is robust to different GNN encoders.

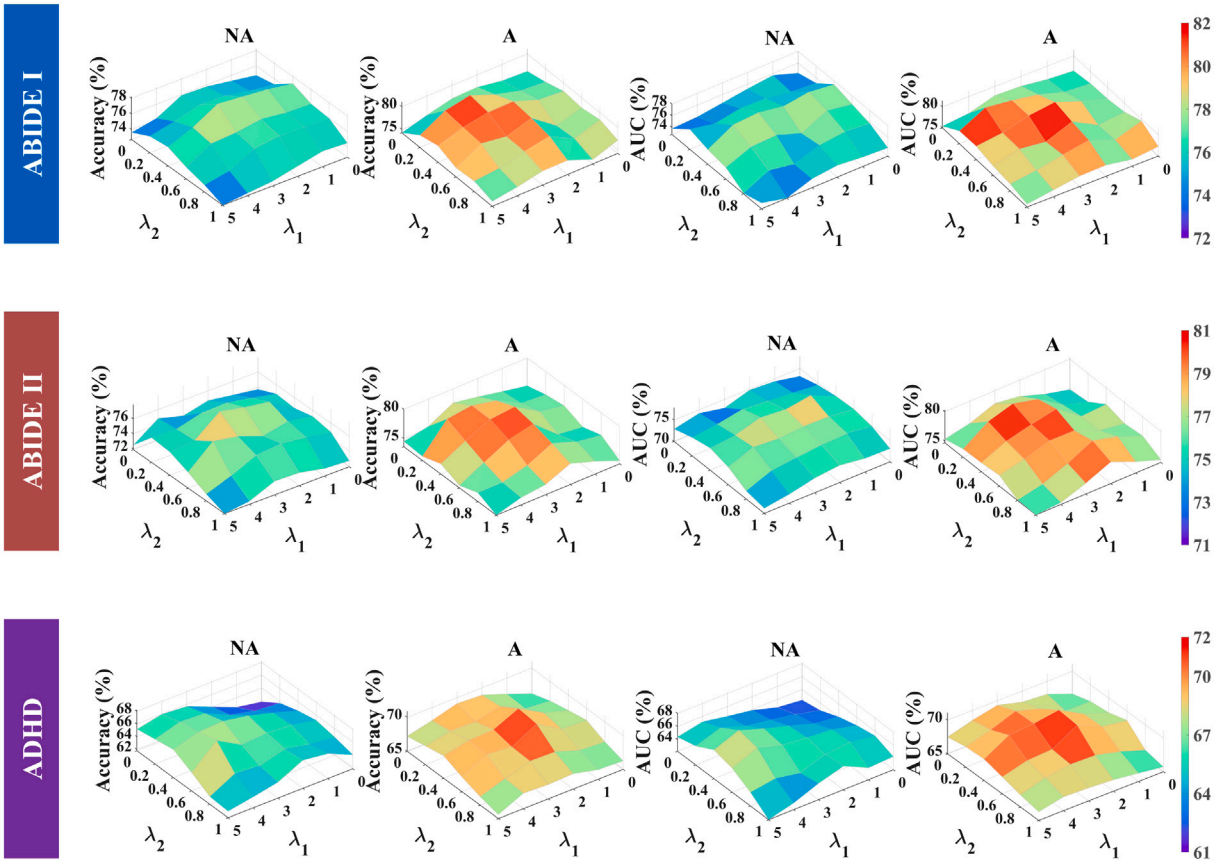


Fig. 6. Ablation study on how λ_1 and λ_2 influence the classification performance given the max–min (adversarial, abbreviated as A) loss function in Eq. (2) and the one-stage (non-adversarial, NA) loss function in Eq. (3). For each paired λ_1 and λ_2, the classification performance in terms of mean accuracy and AUC is shown. When $\lambda_2 = 0$, it corresponds to removing the memory bank.

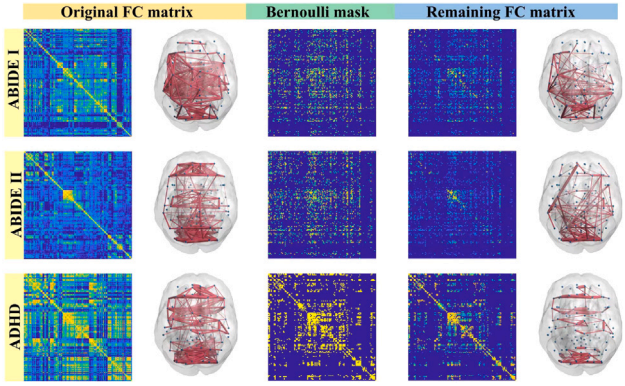


Fig. 7. Illustration of the original FC matrix, the learned Bernoulli mask, and the remaining FC matrix. Only the top 20% FCs are shown.

3.3.4. Influence of the graph augmentation strategy

The traditional contrastive learning approach commonly uses simple data augmentation techniques such as random rotation or intensity scaling in computer vision tasks (Chen et al., 2020). While in graph learning, random edge-dropping is a common method for graph augmentation (Feng et al., 2020). In the proposed A-GCL, we use a more effective Bernoulli mask with adversarial training. To demonstrate the effectiveness of the Bernoulli mask, we compare it with a random edge-dropping strategy. The latter utilizes a random mask to drop edges to generate positive samples, which are fed into the GNN encoder to extract similar latent representations. The results, as shown in Fig. 3(c), demonstrate that the Bernoulli mask generated by A-GCL is more effective than random edge-dropping.

3.3.5. Influence of the embedding dimension

In our experiment, the embedding dimension of latent vectors is set to 32. To investigate the robustness of A-GCL, we change the embedding dimension from 8 to 128. As shown in Fig. 5, the accuracy and AUC increase as the embedding dimension gets larger. Thus, the representation capability of A-GCL generally improves as the embedding dimension increases. Another phenomenon is that 32 may be the optimal embedding dimension, as the performance tends to be stable when the dimension changes from 32 to 128, while larger embedding dimensions require more computational resources.

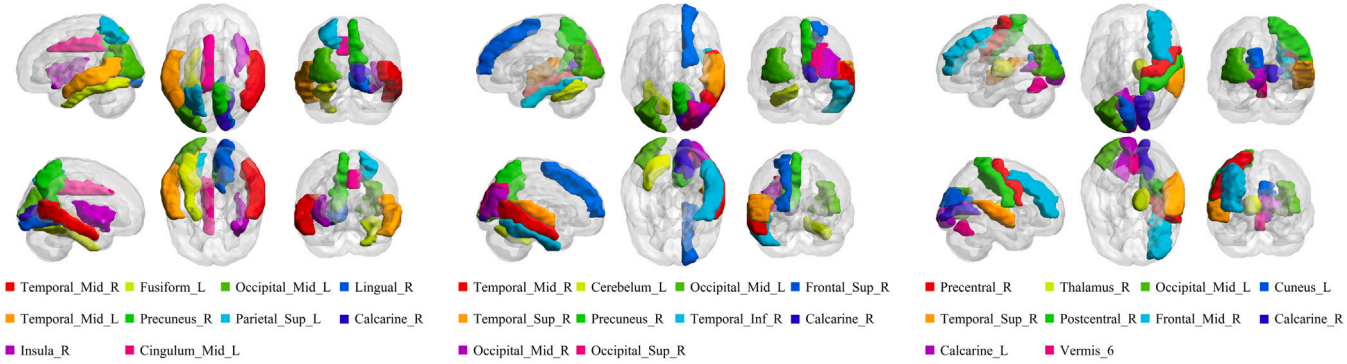
3.3.6. Influence of λ_1, λ_2, and the max–min loss function

To investigate the influence of hyper-parameter λ_1, λ_2 and the max–min loss function, we conduct an ablation study to see how different λ's influence the model performance and whether the max–min (adversarial) training loss function is effective. For comparison, we adopt the following one-stage (non-adversarial) loss function to replace Eq. (2):

$$\min_{z,\mu} -I(z,\mu;\mathcal{B}) + \lambda_1 R(\mu;\mathcal{B}) - \lambda_2 I(z,\mu;\mathcal{B},\mathcal{M}). \tag{3}$$

AAL1 is selected as the atlas in this experiment. As shown in Fig. 6, the best performance is achieved when $\lambda_1 = 2$ and $\lambda_2 = 0.4$ with the max–min (adversarial) loss function.

As shown in Fig. 6, all adversarial versions achieve higher scores than the non-adversarial ones, suggesting that the adversarial loss is more effective than the non-adversarial loss for A-GCL. The better performance from the adversarial training strategy demonstrates the effectiveness of this strategy and the necessity of adversarial edge-dropping during the self-supervised learning process.



(a) ABIDE I (b) ABIDE II (c) ADHD

Fig. 8. Top 10 important brain regions associated with ASD/ADHD based on the three datasets.

3.4. Interpretation

3.4.1. Visualization of the learned Bernoulli mask

To interpret the edge dropping caused by the Bernoulli mask used in A-GCL, we visualize the mean Bernoulli mask across instances in a batch. This allows us to better understand the remaining edges that are related to the changes induced by the disease.

In Fig. 7, we present the original FC matrices, the Bernoulli masks, the edge-dropped FC matrices, and their corresponding brain FC maps, based on AAL1. Only the top 20% connections of the brain FC maps are shown, as the complete FC maps are too dense. Here, the Bernoulli masks are binary (0 or 1) and are used to create the sparse FC matrix via an element-wise multiplication. From Fig. 7, we could observe that the remaining FC maps from the two ABIDE datasets are similar while those from the ADHD-200 dataset have quite different connections. We measure this similarity by calculating the correlation coefficient of the edge-dropped FC matrices between two datasets. It turns out that this correlation coefficient is 0.8328 between ABIDE I and ABIDE II, in contrast to 0.1422/0.1758 between ABIDE I/II and ADHD. This suggests that the disease-related FCs are stable across datasets and may truly reflect the disease-associated changes.

3.4.2. Visualization of the important brain regions

We further analyze the important brain regions associated with ASD or ADHD using A-GCL. The importance of a brain region is quantified by the summed FCs of a node in the edge-dropped graph, i.e., $(A \circ B \circ E)\mathbf{1}$. For each dataset, we calculate this importance score for each brain region, sort the scores, and pick the top 10 most important regions. The results are visualized by BrainNet-Viewer (www.nitrc.org) and shown in Fig. 8.

As shown in Fig. 8, Temporal_Mid_R, Precuneus_R, Occipital_Mid_R, and Calcarine_R are the most important ROIs for ASD classification. Note that these important ROIs from ABIDE I and ABIDE II have also been reported by previous studies (Eilam-Stock et al., 2014). Additionally, the important regions for ADHD – Precentral_R, Thalamus_R, and Frontal_Mid_R – have been found to be highly related to ADHD diagnosis, as reported in several clinical studies (Seidman et al., 2005; Konrad and Eickhoff, 2010).

4. Discussion and conclusion

4.1. Impact of atlas selection

Regarding brain atlas selection, two AAL atlases, AAL1 and AAL3, achieve similar numerical performance for rs-fMRI classification. However, compared to the two AAL atlases, the Shen268 atlas results in lower numerical performance. The possible reason is that the Shen268 atlas divides the brain into more granular regions, which leads to: (1) fewer voxels are averaged, resulting in reduced noise suppression, (2) more nodes exist in the graph, resulting in a larger parameter space and a more difficult optimization problem. An interesting direction in the future is to combine multiple atlases to boost the performance.

4.2. Transfer learning between the two ABIDE datasets

The fact that transfer learning between the two ABIDE datasets is not as effective as training from scratch may seem counter-intuitive. However, note that there is significant heterogeneity between the two ABIDE datasets (Di Martino et al., 2017). They contain raw rs-fMRI data acquired from different MR scanners in 17 sites. In addition, the quality control measures applied to the data are not consistent between these two datasets, resulting in significant differences between them. Therefore, the parameters trained on one dataset may not be suitable for the other, and direct transfer learning may not be appropriate in this case.

4.3. Conclusion

In this paper, we propose A-GCL to diagnose neurodevelopmental disorders such as ASD and ADHD using three rs-fMRI datasets: ABIDE I, ABIDE II, and ADHD-200. A-GCL leverages ALFFs extracted from BOLD signals to build the graph and extracts graph representations by creating a sparsely connected graph as a positive sample in contrastive learning. This enables A-GCL to better aggregate the neighborhood node features in existing GNN-based models. To further enhance the representation capability during the self-supervised process, we implement A-GCL with a dynamic memory bank. The proposed A-GCL is trained by an adversarial strategy. Extensive experiment results demonstrate that A-GCL outperforms other methods in terms of accuracy and other metrics. Additionally, several ablation studies with different atlases verify the stability and robustness of A-GCL, and explanatory experiments provide the disease-associated brain regions, which may benefit both its clinical application and further understanding of the disease.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: One of the co-authors, Dinggang Shen, is affiliated with Shanghai United Imaging Intelligence Co., Ltd. He has declared financial interests. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This study was supported in part by grants from the National Natural Science Foundation of China (82171903, 92043301), the Shanghai Science and Technology Committee (20ZR1407800), and the Shanghai Municipal Science and Technology Major Project (2018SHZDZX01).

References

Almuqhim, F., Saeed, F., 2021. ASD-SAENet: a sparse autoencoder, and deep-neural network model for detecting autism spectrum disorder (ASD) using fMRI data. Front. Comput. Neurosci. 15, 27.

Bessadok, A., Mahjoub, M.A., Rekik, I., 2022. Graph neural networks in network neuroscience. IEEE Trans. Pattern Anal. Mach. Intell..

Bu, X., Hu, X., Zhang, L., Li, B., Zhou, M., Lu, L., Hu, X., Li, H., Yang, Y., Tang, W., et al., 2019. Investigating the predictive value of different resting-state functional MRI parameters in obsessive-compulsive disorder. Transl. Psychiatry 9 (1), 1–10.

Canario, E., Chen, D., Biswal, B., 2021. A review of resting-state fMRI and its use to examine psychiatric disorders. Psychoradiology 1 (1), 42–53.

Cao, M., Yang, M., Qin, C., Zhu, X., Chen, Y., Wang, J., Liu, T., 2021. Using DeepGCN to identify the autism spectrum disorder from multi-site resting-state data. Biomed. Signal Process. Control 70, 103015.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. PMLR, pp. 1597–1607.

Chen, Y., Yan, J., Jiang, M., Zhang, T., Zhao, Z., Zhao, W., Zheng, J., Yao, D., Zhang, R., Kendrick, K.M., Jiang, X., 2022. Adversarial learning based node-edge graph attention networks for autism spectrum disorder identification. IEEE Trans. Neural Netw. Learn. Syst..

Chen, X., Zhang, H., Zhang, L., Shen, C., Lee, S.-W., Shen, D., 2017. Extraction of dynamic functional connectivity from brain grey matter and white matter for MCI classification. Hum. Brain Map. 38 (10), 5019–5034.

Chen, X., Zhou, J., Ke, P., Huang, J., Xiong, D., Huang, Y., Ma, G., Ning, Y., Wu, F., Wu, K., 2023. Classification of schizophrenia patients using a graph convolutional network: A combined functional MRI and connectomics analysis. Biomed. Signal Process. Control 80, 104293.

Chong, C.D., Schwedt, T.J., Hougaard, A., 2019. Brain functional connectivity in headache disorders: a narrative review of MRI investigations. J. Cereb. Blood Flow Metab. 39 (4), 650–669.

Cordova, M., Shada, K., Demeter, D.V., Doyle, O., Miranda-Dominguez, O., Perrone, A., Schifsky, E., Graham, A., Fombonne, E., Langhorst, B., et al., 2020. Heterogeneity of executive function revealed by a functional random forest approach across ADHD and ASD. NeuroImage: Clin. 26, 102245.

Cui, H., Dai, W., Zhu, Y., Kan, X., Gu, A.A.C., Lukemire, J., Zhan, L., He, L., Guo, Y., Yang, C., 2022a. BrainGB: a benchmark for brain network analysis with graph neural networks. IEEE Trans. Med. Imaging.

Cui, H., Dai, W., Zhu, Y., Li, X., He, L., Yang, C., 2022b. Interpretable graph neural networks for connectome-based brain disorder analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 375–385.

Di, X., Biswal, B.B., 2020. Intersubject consistent dynamic connectivity during natural vision revealed by functional MRI. NeuroImage 216, 116698.

Di Martino, A., O'connor, D., Chen, B., Alaerts, K., Anderson, J.S., Assaf, M., Balsters, J.H., Baxter, L., Beggiato, A., Bernaerts, S., et al., 2017. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. Sci. Data 4 (1), 1–15.

Eilam-Stock, T., Xu, P., Cao, M., Gu, X., Van Dam, N.T., Anagnostou, E., Kolevzon, A., Soorya, L., Park, Y., Siller, M., et al., 2014. Abnormal autonomic and associated brain activities during rest in autism spectrum disorder. Brain 137 (1), 153–171.

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. Front. Neuroinform. 13, 70.

Eslami, T., Saeed, F., 2019. Auto-ASD-network: a technique based on deep learning and support vector machines for diagnosing autism spectrum disorder using fMRI data. In: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. pp. 646–651.

Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., et al., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods 16 (1), 111–116.

Feng, W., Zhang, J., Dong, Y., Han, Y., Luan, H., Xu, Q., Yang, Q., Kharlamov, E., Tang, J., 2020. Graph random neural networks for semi-supervised learning on graphs. Adv. Neural Inf. Process. Syst. 33, 22092–22103.

Guo, X., Chen, H., Long, Z., Duan, X., Zhang, Y., Chen, H., 2017. Atypical developmental trajectory of local spontaneous brain activity in autism spectrum disorder. Sci. Rep. 7 (1), 1–10.

Hamilton, W., Ying, Z., Leskovec, J., 2017. Inductive representation learning on large graphs. Adv. Neural Inf. Process. Syst. 30.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R., 2022. Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16000–16009.

He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR.

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: Clin. 17, 16–23.

Hossain, M.D., Kabir, M.A., Anwar, A., Islam, M.Z., 2021. Detecting autism spectrum disorder using machine learning techniques. Health Inf. Sci. Syst. 9 (1), 1–13.

Hull, J.V., Dokovna, L.B., Jacokes, Z.J., Torgerson, C.M., Irimia, A., Van Horn, J.D., 2017. Resting-state functional connectivity in autism spectrum disorders: A review. Front. Psychiatry 7, 205.

Jiang, H., Cao, P., Xu, M., Yang, J., Zaiane, O., 2020. Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. Comput. Biol. Med. 127, 104096.

Kam, T.-E., Suk, H.-I., Lee, S.-W., 2017. Multiple functional networks modeling for autism spectrum disorder diagnosis. Hum. Brain Map. 38 (11), 5804–5821.

Kazi, A., Shekarforoush, S., Arvind Krishna, S., Burwinkel, H., Vivar, G., Kortüm, K., Ahmadi, S.-A., Albarqouni, S., Navab, N., 2019. InceptionGCN: receptive field aware graph convolutional network for disease prediction. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 73–85.

Kim, M., Tack, J., Hwang, S.J., 2020. Adversarial self-supervised contrastive learning. Adv. Neural Inf. Process. Syst. 33, 2983–2994.

Kim, B.-H., Ye, J.C., Kim, J.-J., 2021. Learning dynamic graph representation of brain connectome with spatio-temporal attention. Adv. Neural Inf. Process. Syst. 34, 4314–4327.

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

Konrad, K., Eickhoff, S.B., 2010. Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder. Hum. Brain Map. 31 (6), 904–916.

Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L.H., Ventola, P., Duncan, J.S., 2021. Braingnn: Interpretable brain graph neural network for fmri analysis. Med. Image Anal. 74, 102233.

Li, X., Zhou, Y., Dvornek, N.C., Zhang, M., Zhuang, J., Ventola, P., Duncan, J.S., 2020. Pooling regularized graph neural network for fmri biomarker analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 625–635.

Liu, Y., Xu, L., Li, J., Yu, J., Yu, X., 2020. Attentional connectivity-based prediction of autism using heterogeneous rs-fMRI data from CC200 atlas. Exp. Neurobiol. 29 (1), 27.

Luo, D., Cheng, W., Xu, D., Yu, W., Zong, B., Chen, H., Zhang, X., 2020. Parameterized explainer for graph neural network. Adv. Neural Inf. Process. Syst. 33, 19620–19631.

Parisot, S., 2018. Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. Med. Image Anal. 48, 117–130.

Peng, L., Wang, N., Xu, J., Zhu, X., Li, X., 2022. GATE: Graph CCA for temporal self-supervised learning for label-efficient fMRI analysis. IEEE Trans. Med. Imaging.

Power, J.D., 2017. A simple but useful way to assess fMRI scan qualities. Neuroimage 154, 150–158.

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59 (3), 2142–2154.

Rolls, E.T., Huang, C.-C., Lin, C.-P., Feng, J., Joliot, M., 2020. Automated anatomical labelling atlas 3. Neuroimage 206, 116189.

Seidman, L.J., Valera, E.M., Makris, N., 2005. Structural brain imaging of attention-deficit/hyperactivity disorder. Biol. Psychiatry 57 (11), 1263–1272.

Shen, X., Tokoglu, F., Papademetris, X., Constable, R.T., 2013. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. Neuroimage 82, 403–415.

Soon, C.S., Vinogradova, K., Ong, J.L., Calhoun, V.D., Liu, T., Zhou, J.H., Ng, K.K., Chee, M.W., 2021. Respiratory, cardiac, EEG, BOLD signals and functional connectivity over multiple microsleep episodes. NeuroImage 237, 118129.

Sun, L., Yu, K., Batmanghelich, K., 2021. Context matters: Graph-based self-supervised representation learning for medical images. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35. pp. 4874–4882.

Suresh, S., Li, P., Hao, C., Neville, J., 2021. Adversarial graph augmentation to improve graph contrastive learning. Adv. Neural Inf. Process. Syst. 34.

Tang, H., Ma, G., Guo, L., Fu, X., Huang, H., Zhan, L., 2022. Contrastive brain network learning via hierarchical signed graph pooling model. IEEE Trans. Neural Netw. Learn. Syst..

Veličković, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D., 2018. Deep graph infomax. Int. Conf. Learn. Represent. 2 (3), 1–17.

Wang, X., Yao, L., Rekik, I., Zhang, Y., 2022b. Contrastive functional connectivity graph learning for population-based fMRI classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 221–230.

Wang, J., Zhang, F., Jia, X., Wang, X., Zhang, H., Ying, S., Wang, Q., Shi, J., Shen, D., 2022a. Multi-class ASD classification via label distribution learning with class-shared and class-specific decomposition. Med. Image Anal. 75, 102294.

Xu, D., Cheng, W., Luo, D., Chen, H., Zhang, X., 2021. Infogcl: Information-aware graph contrastive learning. Adv. Neural Inf. Process. Syst. 34, 30414–30425.

Xu, K., Hu, W., Leskovec, J., Jegelka, S., 2018. How powerful are graph neural networks? arXiv preprint arXiv:1810.00826.

Yan, Y., Dahmani, L., Ren, J., Shen, L., Peng, X., Wang, R., He, C., Jiang, C., Gong, C., Tian, Y., et al., 2020. Reconstructing lost BOLD signal in individual participants using deep machine learning. Nat. Commun. 11 (1), 1–13.

Yang, Y., Cui, H., Yang, C., 2023. PTGB: Pre-train graph neural networks for brain network analysis. In: Conference on Health, Inference, and Learning. PMLR, pp. 526–544.

Yang, H., Long, X.-Y., Yang, Y., Yan, H., Zhu, C.-Z., Zhou, X.-P., Zang, Y.-F., Gong, Q.-Y., 2007. Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI. Neuroimage 36 (1), 144–152.

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.

You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., Shen, Y., 2020. Graph contrastive learning with augmentations. Adv. Neural Inf. Process. Syst. 33, 5812–5823.

Zhang, H., Song, R., Wang, L., Zhang, L., Wang, D., Wang, C., Zhang, W., 2022. Classification of brain disorders in rs-fMRI via local-to-global graph neural networks. IEEE Trans. Med. Imaging.

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

Zhou, H., Zhang, D., 2021. Graph-in-graph convolutional networks for brain disease diagnosis. In: 2021 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 111–115.