Contrastive Graph Pooling for Explainable Classification of Brain Networks

Jiaxing Xu, Qingtian Bian, Xinhang Li, Aihu Zhang, Yiping Ke, Miao Qiao,
Wei Zhang, Wei Khang Jeremy Sim, and Balázs Gulyás
for the Alzheimer's Disease Neuroimaging Initiative [†]

Abstract—**Functional magnetic resonance imaging (fMRI) is a commonly used technique to measure neural activation. Its application has been particularly important in identifying underlying neurodegenerative conditions such as Parkinson's, Alzheimer's, and Autism. Recent analysis of fMRI data models the brain as a graph and extracts features by graph neural networks (GNNs). However, the unique characteristics of fMRI data require a special design of GNN. Tailoring GNN to generate effective and domain-explainable features remains challenging. In this paper, we propose a contrastive dual-attention block and a differentiable graph pooling method called *ContrastPool* to better utilize GNN for brain networks, meeting fMRI-specific requirements. We apply our method to 5 resting-state fMRI brain network datasets of 3 diseases and demonstrate its superiority over state-of-the-art baselines. Our case study confirms that the patterns extracted by our method match the domain knowledge in neuroscience literature, and disclose direct and interesting insights. Our contributions underscore the potential of ContrastPool for advancing the understanding of brain networks and neurodegenerative conditions. The source code is available at `https://github.com/AngusMonroe/ContrastPool`.**

Jiaxing Xu (corresponding author), Qingtian Bian, Aihu Zhang, and Yiping Ke are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, 639798 Singapore (e-mail:`JIAXING003@e.ntu.edu.sg`; `BIAN0027@e.ntu.edu.sg`; `ZHAN0547@e.ntu.edu.sg`; `ypke@ntu.edu.sg`).

Xinhang Li is with the Department of Computer Science and Technology, Tsinghua University BEIJING, 100084 P.R.China (e-mail: `xh-li20@mails.tsinghua.edu.cn`).

Miao Qiao is with the School of Computer Science, the University of Auckland, Auckland 1142 New Zealand (e-mail: `miao.qiao@auckland.ac.nz`).

Wei Zhang is with the Cognitive Neuroimaging Centre and Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, 636921 Singapore (e-mail: `wilson.zhangwei@ntu.edu.sg`).

Wei Khang Jeremy Sim is with the Cognitive Neuroimaging Centre and IGP-Neuroscience, Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore, 636921 Singapore (e-mail: `SIMW0035@e.ntu.edu.sg`).

Balázs Gulyás is with the Cognitive Neuroimaging Centre and Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, 636921 Singapore, and the Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden (e-mail: `balazs.gulyas@ntu.edu.sg`).

Index Terms—**Brain Network, Deep Learning for Neuroimaging, fMRI Biomarker, Graph Classification, Graph Neural Network.**

I. INTRODUCTION

FUNCTIONAL magnetic resonance imaging (fMRI) measures neural activation and has been widely used in analyzing functional connectivities of the human brain [1]. Particularly, fMRI has been instrumental in identifying underlying neurodegenerative conditions [2], e.g., Alzheimer's, Parkinson's, and Autism.

A line of research in analyzing fMRI data models brains as graphs, i.e., defining brain regions of interest (ROIs) as nodes and the functional connectivities/interactions between ROIs as edges [3]. Graph analytic methods are then applied to such brain networks to understand and analyze the elements and interactions of neurological systems from a network perspective. While previous research has showcased the remarkable potential of network-based approaches [4], [5] in deciphering the complexities of the brain, graph analysis for brain networks is still an emerging field in its early stages of development.

Recently graph neural networks (GNNs) have attracted increasing interest in graph classification and thus have been applied to graph-structured data in a wide range of applications, including social networks [6], molecular data [7], [8] and recommendation systems [9], [10]. Compared with conventional machine learning models that handle vector-based data, GNNs are able to engage graph topological information through message passing. In light of the promising performance of GNNs on other applications, several studies [11]–[13] have applied GNNs to brain network analysis. However, directly applying general-purpose GNNs to brain networks could be ineffective on fMRI data due to its unique data characteristics [14]: (1) **Low signal-to-noise ratio (SNR).** fMRI is a type of high dimensional data where non-neural noise derived from cardiac/respiratory processes or scanner instability could cause large variations within a single subject and across different subjects. (2) **Node alignment.** Different from other graph-based datasets, each node of a brain network represents an ROI under a brain parcellation scheme. If the same scheme is applied to all subjects, each resulting network will have the same number of nodes and the nodes are aligned across different subjects. (3) **Limited data scale.** Due to the limited availability of the fMRI data, brain network datasets

contain a relatively small number of subjects, which can cause overfitting on GNNs.

To better utilize GNNs for brain networks and meet the need for fMRI characteristics, we propose a contrastive graph pooling method (*ContrastPool*) with a contrastive dual-attention to extract group-specific information. In contrastPool, the ROI-wise attention is introduced to identify the most representative brain regions in each class of subjects for the given task (e.g., the Alzheimer's disease classification). This design aims to reduce the noise caused by the fMRI acquisition process by directing the model to focus on just a few ROIs rather than the entire set, which alleviates the issue of low SNR. In addition, ContrastPool devises a subject-wise attention that leverages the node alignment in brain networks. It aligns different subjects in the same class by ROI and assigns higher attention scores to subjects that are more indicative of the class. By performing aggregation within groups and contrast across groups, ContrastPool generates a contrast graph and uses it to guide the message passing in brain network representation learning. The group knowledge can weaken the sensitivity to specific subjects and thus mitigates the overfitting issue. The contrast operation further identifies the discrepancy across groups and strengthens the dual attention to focus on more discriminative ROIs and subjects. Different from graph contrastive learning methods [15]–[17] which construct positive and negative pairs by data augmentation, we make contrast across groups. The rationale behind this design is to identify the discrepant features discovered from different groups. Therefore, we use a contrast graph to extract the most discriminative task-related features from the training set and use it as prior knowledge to guide brain network representation learning. Our key contributions are:

- We propose a contrastive dual-attention block, which adaptively assigns a weight to each ROI of each subject, performs adaptive aggregation over subjects in each group, and makes contrast across different groups to obtain a contrast graph. Guided by the contrast graph, we introduce a differentiable graph pooling method called *ContrastPool* to generate brain network representations that are effective to the task of disease classification.
- We apply our ContrastPool to 5 resting-state fMRI brain network datasets spanning over 3 diseases. The results justify the superiority of ContrastPool over the state-of-the-art baselines.
- The case study confirms the interestingness, simplicity and high explainability of the patterns extracted by our method, which match the domain knowledge in neuroscience literature.

II. RELATED WORK

A. General-purposed GNNs

GNNs with Attention Mechanism. The integration of attention mechanisms in GNNs [18], [19] involves incorporating attention-based mechanisms to selectively weigh the importance of different neighbors during the information aggregation process. This attention mechanism enhances the modeling of relationships between nodes in a graph, allowing the model to focus on more relevant neighbors. The relevancy of a neighbor is decided by the dynamic and automated assignment of the attention weights. Such an assignment is based on the content and features of neighboring nodes, which enables the GNN to capture complex and non-linear dependencies in graph-structured data, and thus leads to enhanced performances in various tasks [20], [21]. GAT [18] first introduces the attention mechanism to GNNs by utilizing the similarity of two end-points to control the weight of the edge for message passing. AGNN [22] removes all intermediate fully connected layers of graph convolutional network and replaces the propagation layer with an attention mechanism that respects the graph structure. GAM [23] proposes structural attention to enhance graph structure learning for graph classification. Different from these works that introduce attentions to neighbors/graph structures and use the attentions directly in message passing, our work imposes attentions on ROIs and subjects in brain network analysis. Our dual attentions are coupled with group aggregation and contrast to produce a contrast graph, which is then used to guide the message passing. In this sense, our attention mechanism is not applied directly to the message-passing process of GNNs.

Graph Contrastive Learning. In typical graph contrastive learning, positive and negative samples are generated by data augmentation on the graph structure or node features [15]. SUGAR [24] generates subgraphs and uses these subgraphs for reconstruction. AutoGCL [16] employs a set of learnable graph view generators orchestrated by an auto augmentation strategy, where every graph view generator learns a probability distribution of graphs conditioned by the input. ASP [17] proposes a novel attribute and structure-preserving graph contrastive learning framework. The objective of these methods is to maximize the similarities of positive pairs and minimize the similarities of negative pairs. On the contrary, our contrastive dual-attention aims to highlight the discrepancy across different groups, which is based on ground-truth labels instead of generated samples.

Graph Pooling. Graph pooling is a technique in graph neural networks that downsamples graphs by aggregating information from groups of nodes, reducing graph size while preserving essential structural and contextual features for efficient and effective learning. Existing graph pooling methods often select nodes or generate super-nodes by using attention mechanisms [25], maximizing mutual information [26] or encoding graph structures [27]. However, these methods consider the feature and structural information of each graph individually without considering the similarity within groups and the discrepancy across groups.

B. Models for Brain Networks

In recent years, several GNN-based methods have been proposed for brain networks. [11] leverages graph convolutional networks (GCNs) for learning similarities between each pair of graphs (subjects). [4] proposes edge-to-edge, edge-to-node and node-to-graph convolutional filters to leverage the topological information of brain networks in the neural network. [28] puts forward a two-stage pipeline to discover

ASD brain biomarkers from task-fMRI using GNNs. [13] proposes an ROI-selection pooling to highlight salient ROIs for each individual. MG2G [29] adopts a two-stage approach that first learns latent distribution-based node representations by an unsupervised stochastic graph embedding model and then employs the learnt representations to train a classifier. The distribution-based nature of node representations enables the identification of significant ROIs with AD-related effects. These studies neglect the characteristics of fMRI data elaborated earlier in their model design. These works neglect the three characteristics of fMRI data elaborated in Section I. Besides, [12] proposes a graph pooling method with group-level regularization to guarantee group-level consistency. [14] jointly learns the node grouping and extracts the graph features. These two methods only take group information into account on graph level without utilizing node alignment. [5] extracts a dense contrast subgraph to filter useful information for prediction. However, their feature extraction and subject classification treat all ROIs and all subjects equally, which could be vulnerable to noisy data. [30] incorporates local ROI-GNN and global subject-GNN guided by non-imaging data, such as gender, age, and acquisition site. The local ROI-GNN does not take the node alignment of brain networks into account. Ding et al. [31] also integrate embeddings from separate graph views of images and non-imaging data by a co-attention module, along with a contrastive loss-based fusion method to improve cross-view fusion learning. Their contrast is applied to the population graphs of two modalities rather than groups, to facilitate effective fusion of information from different data modalities. Moreover, [32] and [33] propose a graph generator to transform the raw blood-oxygen-level-dependent (BOLD) time series into task-aware brain connectivities. They model subjects as dynamic graphs, which require all subjects to have the same scan length and preferably from the same site. In contrast, our work constructs static brain networks as we operate on datasets collected from multi-sites with various acquisition lengths. A Transformer-based method [34] has been applied to brain networks to learn pairwise connection strengths among ROIs across individuals. It neglects the group information of subjects in its methodology design.

III. METHODOLOGY

In this section, we first introduce the construction of brain networks and commonly-used GNN schemes for graph classification. We then elaborate the design of our proposed ContrastPool. Notation-wise, we use calligraphic letters (e.g., $\mathcal{G}$) to denote sets, bold capital letters (e.g., $\boldsymbol{A}$) matrices, and bold lowercase letters (e.g., $\boldsymbol{z}$) vectors. Subscripts and superscripts are used to distinguish between different variables or parameters, while lowercase letters denote scalars. We use $\boldsymbol{A}(i,:)$ and $\boldsymbol{A}(:,j)$ to denote the i-th row and j-th column of a matrix $\boldsymbol{A}$, respectively. This notation also extends to a 3D matrix. Table I summarizes the notations used throughout the paper.

A. Preliminaries

Brain Network Construction. We construct five brain network datasets from five neuro-imaging data sources, which

TABLE I
NOTATION TABLE

Notation	Description
$\boldsymbol{M}$	Connectivity matrix of a subject
G	An input graph/brain network
$\boldsymbol{A}$	Adjacency matrix of G
$\boldsymbol{X}$	Node feature matrix of G
$\mathcal{V}_G$	Node set of G
v, u	A node in G
$\boldsymbol{H}$	Node representations
$\boldsymbol{H}_v$	Node representation of v
m	Number of nodes in G
$\mathcal{D}$	Input dataset
$\mathcal{G}$	Input graph set
$\mathcal{Y}$	Input label set
y_G	Label of G
l	The layer index in GNN
d	Dimensionality of node representation $\boldsymbol{H}_v$
$\mathcal{G}^{TC}, \mathcal{G}^{ASD}$	Graph set of TC/ASD group
$\mathcal{A}^{TC}, \mathcal{A}^{ASD}$	Adjacency matrix set of TC/ASD group
$\mathcal{X}^{TC}, \mathcal{X}^{ASD}$	Feature matrix set of TC/ASD group
$G_{sum}^{TC}, G_{sum}^{ASD}$	Summary graph of TC/ASD group
$\boldsymbol{A}_{sum}^{TC}, \boldsymbol{A}_{sum}^{ASD}$	Summary adjacency matrix of TC/ASD group
$\boldsymbol{X}_{sum}^{TC}, \boldsymbol{X}_{sum}^{ASD}$	Summary feature matrix of TC/ASD group
$G_{contrast}$	Contrast graph
$\boldsymbol{A}_{contrast}$	Adjacency matrix of $G_{contrast}$
$\boldsymbol{H}_{contrast}$	Node representations of $G_{contrast}$
$\boldsymbol{W}_Q, \boldsymbol{W}_K, \boldsymbol{W}_V$	Parameter matrix
i, j	Index for matrix dimensions
n^{TC}	Number of subjects in TC group
$\boldsymbol{T}_{ROI}^{TC}$	Output 3D matrix of $\text{Attn}_{ROI}(\cdot)$ for TC group
$\boldsymbol{T}_{subject}^{TC}$	Output 3D matrix of $\text{Attn}_{subject}(\cdot)$ for TC group
$\boldsymbol{S}$	Cluster assignment matrix
$\boldsymbol{Z}$	Embedded node feature matrix
$\boldsymbol{A}_{contrast}(i,:)$	The i-th row of $\boldsymbol{A}_{contrast}$

are Taowu, PPMI and Neurocon [35] for Parkinson's disease, ADNI [36] for Alzheimer's disease, and ABIDE [37] for Autism.

[TaoWu and Neurocon] The TaoWu and Neurocon datasets, released by ICI [35], are among the earliest image datasets made available for Parkinson's research. These datasets comprise age-matched subjects collected from a single machine and site, encompassing both normal control individuals and patients diagnosed with PD.

[PPMI] The Parkinson's Progression Markers Initiative (PPMI) is a comprehensive study aiming to identify biological markers associated with Parkinson's risk, onset, and progression. PPMI comprises multimodal and multi-site MRI images. The dataset consists of subjects from 4 distinct classes: normal control (NC), scans without evidence of dopaminergic deficit (SWEDD), prodromal, and Parkinson's disease (PD). This ongoing study plays a crucial role in advancing our understanding of Parkinson's disease.

[ADNI] The ADNI raw images used in this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Prin-

cipal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see `www.adni-info.org`. We included subjects from 6 different stages of AD, from cognitive normal (CN), significant memory concern (SMC), mild cognitive impairment (MCI), early MCI (EMCI), late MCI (LMCI) to Alzheimer's disease (AD).

[ABIDE] The Autism Brain Imaging Data Exchange (ABIDE) initiative supports the research on Autism Spectrum Disorder (ASD) by aggregating functional brain imaging data from laboratories worldwide. ASD is characterized by stereotyped behaviors, including irritability, hyperactivity, depression, and anxiety. Subjects in the dataset are classified into two groups: typical controls (TC) and individuals diagnosed with ASD.

TABLE II

STATISTICS OF BRAIN NETWORK DATASETS.

Dataset	Condition	Graph#	Avg BOLD length	Class#
Taowu	Parkinson	40	239	2
PPMI	Parkinson	209	198	4
Neurocon	Parkinson	41	137	2
ADNI	Alzheimer	1326	344	6
ABIDE	Autism	989	201	2

We followed [3] in resting-state fMRI data preprocessing and brain network construction. Specifically, raw MRI images are collected from the five data sources. Each fMRI image is accompanied with a structural T1-weighted (T1w) image acquired from the same subject in the same scan session. The fMRI images are then preprocessed using fMRIPrep [38], with T1w used to extract the brain mask and perform image alignment and normalization of BOLD signals. The preprocessing involves a number of pipeline steps, including motion correction, realigning, field unwarping, normalization, bias field correction, and brain extraction. Schaefer atlas [39] is applied to the preprocessed images to parcellate the brain into 100 ROIs. A brain network is constructed for each subject, which is represented by a connectivity matrix M. The nodes in M are ROIs and the edges are the Pearson's correlation between the region-averaged BOLD signals from pairs of ROIs. Essentially, M captures functional relationships between different ROIs. Statistics of the brain network datasets are summarized in Table II. The class/group distribution of the datasets is provided in Table III.

Brain Network Classification. Generally, graph classification aims to predict certain properties (i.e., class labels) of given graphs. Given a dataset of labeled graphs $\mathcal{D} = (\mathcal{G}, \mathcal{Y}) = \{(G, y_G)\}$, where y_G is the class label of a graph $G \in \mathcal{G}$, the problem of graph classification is to learn a predictive function $f: \mathcal{G} \to \mathcal{Y}$, which maps input graphs to their labels. Take brain networks constructed from the Autism dataset as an example. The class label set contains two labels: TC and ASD. The objective of brain network classification is to learn a predictive function f from a set of brain networks, expecting that f also works well on unseen brain networks.

TABLE III

CLASS DISTRIBUTION OF BRAIN NETWORK DATASETS.

Dataset	Class	# Subjects
ABIDE	Control	534
	ASD	455
ADNI	CN	819
	SMC	72
	LMCI	102
	EMCI	89
	MCI	179
	AD	65
PPMI	Control	15
	SWEDD	14
	Prodromal	67
	PD	113
TaoWu	Control	20
	PD	20
Neurocon	Control	15
	PD	26

Graph Neural Networks. Compared with conventional vector-based machine learning models, GNNs engage graph topological information in graph representation learning. Consider an input graph $G = (A, X)$, where A is the adjacency matrix, which encodes connectivity information, and X is the feature matrix containing attribute information for each node. The node set of G is denoted as $\mathcal{V}_G$ and $|\mathcal{V}_G| = m$. The l-th layer of a GNN in the message-passing scheme [40] can be written as:

$$\boldsymbol{H}_v^{(l)} = \text{AGG}^{(l-1)}\left(\boldsymbol{H}_v^{(l-1)}, \text{MSG}^{(l-1)}\left(\left\{\boldsymbol{H}_u^{(l-1)}\right\}_{\forall u \in \mathcal{N}(v)}\right)\right). \tag{1}$$

Herein, $\boldsymbol{H}^{(l)} \in \mathbb{R}^{m \times d}$ denotes the l-th layer node representation, where each node is represented by a d-dimensional vector. $\text{AGG}(\cdot)$ and $\text{MSG}(\cdot)$ are arbitrary differentiable aggregate and message functions (e.g., a multilayer perceptron (MLP) can be used as $\text{AGG}(\cdot)$ and a summation function as $\text{MSG}(\cdot)$). $\mathcal{N}(v)$ represents the neighbor node set of node $v \in \mathcal{V}_G$, and $\boldsymbol{H}_v^{(0)} = \boldsymbol{X}_v$. The updated representations then pass through a sum/mean pooling and are fed to a linear layer for classification. In brain network classification, we use the connectivity matrix M as both the adjacency matrix A and feature matrix X. Despite prevalent, GNNs implementing Eq. (1) only propagate information across the edges of each individual brain network, and thus neglect the group-based information and node alignment. This standard message passing scheme is also vulnerable to noise as it treats all ROIs and all subjects equally important in its representation learning process. Therefore, the goal of this work is to design an end-to-end framework that addresses these limitations of GNNs when applied to brain networks.

B. Contrastive Graph Pooling with Dual-Attention

We propose ContrastPool, which addresses the above challenges by aggregating subjects over different classes using dual-attention. The idea of ContrastPool is based on an important observation: the relatedness of different ROIs for different diseases varies, and the extent to which a subject exhibits typical characteristics of a disease also varies. Therefore, we design ContrastPool in a way that the graphs in the
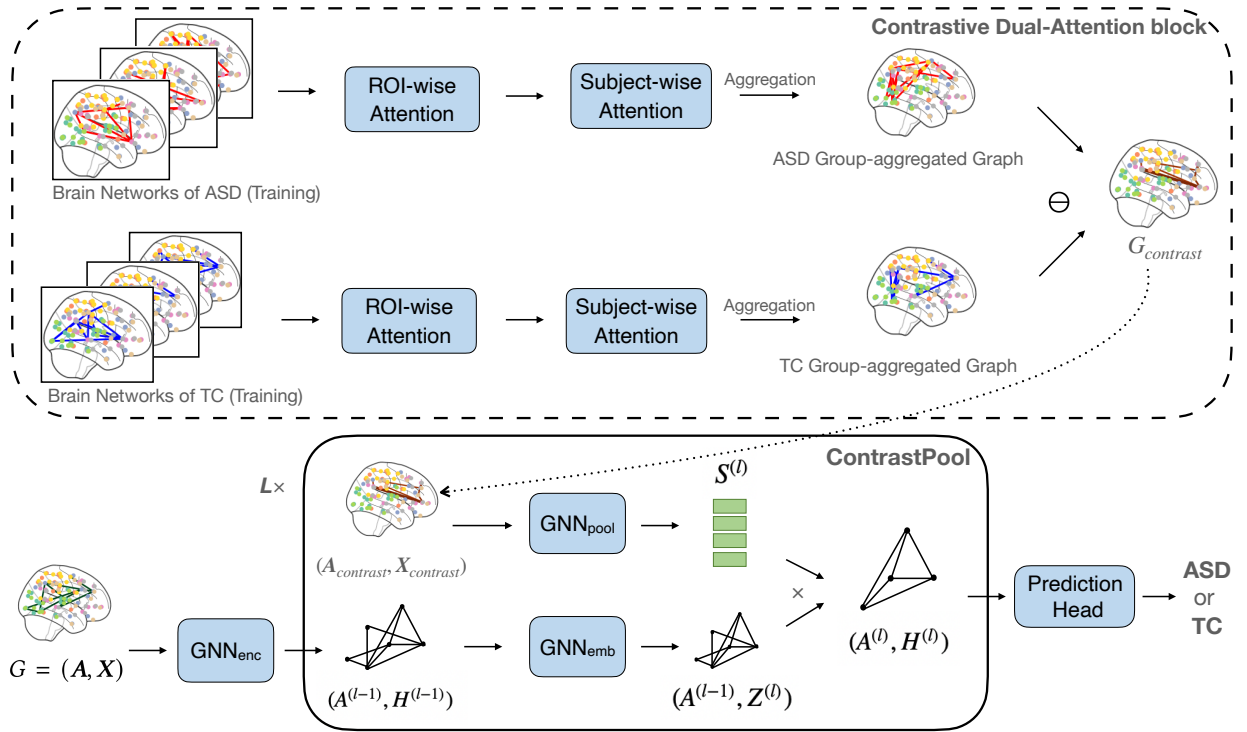
Fig. 1. The architecture of ContrastPool, using Autism as an example.

same group (e.g., ASD group) are summarized by assigning different weights for ROIs and subjects. These weights are not assigned manually but learnt automatically from training data to optimize the classification performance. The architecture of ContrastPool is shown in Fig. 1, with Autism as an example. In the training stage, all training graphs are categorized by their groups and pass through the Contrastive Dual-Attention (CDA) block to generate the contrast graph (the upper block of Fig. 1). This contrast graph is subsequently encoded into an assignment matrix, which is combined with each input graph for hierarchical graph pooling (the lower block of Fig. 1). In the validation and test stages, each validation/test graph passes through the lower block and is fused with the contrast graph (generated by the training phase) to obtain its graph representation. This representation is then fed into the Prediction Head to produce the final prediction of the graph (either ASD or TC in the example).

In the following, we first introduce the Contrastive Dual-Attention block, which is used to generate a contrast graph. We then describe how the ContrastPool module produces node and graph representations based on the contrast graph. Finally, we discuss our design of the loss function.

Contrastive Dual-Attention (CDA) Block. This block aims to generate a contrast graph that best characterizes the differences in brain networks between two groups of subjects. In order to achieve this, it first summarizes all the graphs within the same group into a summary graph. It then computes the difference of the summary graphs from two different groups. We design the computation between the summary graph as a learnable dual-attention process such that the most discriminative ROIs and the most representative subjects can be automatically

highlighted in the summary graph formation.

Consider two groups of graphs in the training set, $\mathcal{G}^{TC} = (\mathcal{A}^{TC}, \mathcal{X}^{TC})$ and $\mathcal{G}^{ASD} = (\mathcal{A}^{ASD}, \mathcal{X}^{ASD})$, bearing TC and ASD groups, respectively. CDA first computes the summary graphs G_{sum}^{TC} and G_{sum}^{ASD} of the two groups via two subject aggregation functions $\mathrm{SA}_A(\cdot)$ and $\mathrm{SA}_X(\cdot)$ performed respectively on the adjacency matrix set $\mathcal{A}^{TC}$, $\mathcal{A}^{ASD}$ and the feature matrix set $\mathcal{X}^{TC}$, $\mathcal{X}^{ASD}$:

$$G_{sum}^{TC} = (\boldsymbol{A}_{sum}^{TC}, \boldsymbol{X}_{sum}^{TC}) = (\mathrm{SA}_A(\mathcal{A}^{TC}), \mathrm{SA}_X(\mathcal{X}^{TC})), \quad (2)$$

$$G_{sum}^{ASD} = (\boldsymbol{A}_{sum}^{ASD}, \boldsymbol{X}_{sum}^{ASD}) = (\mathrm{SA}_A(\mathcal{A}^{ASD}), \mathrm{SA}_X(\mathcal{X}^{ASD})). \quad (3)$$

The contrast graph $G_{contrast} = (\boldsymbol{A}_{contrast}, \boldsymbol{H}_{contrast})$ is then obtained by:

$$\boldsymbol{A}_{contrast} = \boldsymbol{A}_{sum}^{TC} \ominus \boldsymbol{A}_{sum}^{ASD}, \quad (4)$$

$$\boldsymbol{H}_{contrast} = \boldsymbol{X}_{sum}^{TC} \ominus \boldsymbol{X}_{sum}^{ASD}, \quad (5)$$

where $\ominus$ is a binary function that computes the element-wise absolute differences of two matrices.

To enable the automatic learning of the weights on the ROIs and subjects, respectively, we design a dual-attention function as $\mathrm{SA}(\cdot)$. In general, a self-attention function [41] for a given matrix $\boldsymbol{X} \in \mathbb{R}^{k \times m}$ can be written as:

$$\mathrm{Attn}(\boldsymbol{X}) = \mathrm{norm}\left(\boldsymbol{X} + \phi\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{k}}\right)\boldsymbol{V}\right), \quad (6)$$

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}_Q, \boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}_K, \boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}_V, \qquad (7)$$

where $\boldsymbol{W}_Q, \boldsymbol{W}_K, \boldsymbol{W}_V \in \mathbb{R}^{m \times m}$ are parameter matrices, $\phi(\boldsymbol{z})_i = \mathrm{softmax}(\boldsymbol{z})_i = \frac{e^{\boldsymbol{z}_i}}{\sum_{j=1}^{m} e^{\boldsymbol{z}_j}}, \text{for } i = 1 \dots k, \boldsymbol{z} \in \mathbb{R}^k,$ and $\mathrm{norm}(\cdot)$ is layer normalization function. Essentially, the dual-attention mechanism is performed by feeding different input matrices to Eq. (6) for different levels of attention. We illustrate the dual-attention process with $\mathcal{X}^{TC}$ as an example below. We first apply an ROI-wise attention to compute the attention between the ROIs in each subject. That is, the input matrix for the ROI-attention function $\mathrm{Attn}_{ROI}(\cdot)$ is the adjacency matrix of each subject with $k = m$. On top of this ROI-wise attention, we apply subject-wise attention to compute the attention between subjects in each ROI. The input matrix for the subject-wise attention function $\mathrm{Attn}_{subject}(\cdot)$ is the subject-stacked resultant matrix of each ROI, with a dimension of $n^{TC} \times d$, where n^{TC} is the number of subjects in $\mathcal{X}^{TC}$. Finally, the output of the subject-wise attention is averaged on the subject dimension to obtain the summary feature matrix. The aggregation function $\mathrm{SA}_X(\cdot)$ performed on $\mathcal{X}^{TC}$ can be formally written as:

$$\boldsymbol{T}_{ROI}^{TC} = \left[\mathrm{Attn}_{ROI}\left(\boldsymbol{X}\right) : \boldsymbol{X} \in \mathcal{X}^{TC} \right], \qquad (8)$$

$$\boldsymbol{T}_{subject}^{TC} = \left[\mathrm{Attn}_{subject}\left(\boldsymbol{T}_{ROI}^{TC}(:, i, :)\right) : i = 1, \dots, m \right], \quad (9)$$

$$\mathrm{SA}_X(\mathcal{X}^{TC}) = \frac{\sum_{j=1}^{n^{TC}} \boldsymbol{T}_{subject}^{TC}(j, :, :)}{n^{TC}}, \qquad (10)$$

where $\boldsymbol{T}_{ROI}^{TC}, \boldsymbol{T}_{subject}^{TC} \in \mathbb{R}^{n^{TC} \times m \times m}$ denote two 3-dimensional matrices to store the outputs of the ROI-wise and subject-wise attention functions, and $[\cdot]$ denotes the stack operation to combine a set of 2D matrices to a single 3D matrix. The subject aggregation function $\mathrm{SA}_A(\cdot)$ performed on the adjacency matrix sets can be defined in a similar way.

The ROI-wise attention in Eq. (8) aims to extract discriminative-related ROIs and to mitigate the influence of noise (caused by factors such as cardiac/respiratory noise or scanner instability). We do an empirical study of how ROI-wise attention reduces site-specific noise in Section IV-D. The goal of subject-wise attention in Eq. (9) is to leverage the node alignment and focus on the most representative subjects in each class. Utilizing ROI- and subject-wise attention to filter crucial information from two different aspects enables us to extract features more effectively from brain networks. This design of dual-attention can also guarantee the permutation invariance of brain networks, which means no matter how we permute the indices of subjects or ROIs, the output will be the same. By averaging over all subjects from the same group to compute the summary graph via Eq. (10), the impact of the number of subjects in different groups is alleviated. This helps mitigate the class imbalance issue in some datasets (e.g., PPMI and ADNI). The contrast between summary graphs can also weaken the sensitivity to the non-representative subjects and thus mitigate overfitting.

Pooling with a Contrast Graph. This module aims to produce a high-quality representation for each input graph by leveraging the group-discriminative information captured in the contrast graph. To achieve a natural and effective utilization of the contrast graph in graph pooling and meanwhile extract high-level node features, we adopt DiffPool [42] as the pooling method. The idea is to coarsen the input graph in a hierarchical manner (via layers) such that similar nodes are grouped together into clusters at each layer to extract high-level node representations. Specifically, our ContrastPool takes the input graph with m number of nodes and coarsens it by grouping the input nodes into e.g., $\frac{m}{2}$ number of clusters in a soft manner. The node grouping is performed on an embedded node feature matrix $\boldsymbol{Z}$ and guided by a cluster assignment matrix $\boldsymbol{S}$ that characterizes the node similarity. The output coarsened graph after this pooling layer contains $\frac{m}{2}$ number of nodes, with each node representing a soft cluster of the input nodes. For hierarchical pooling, L pooling layers can be deployed.

In ContrastPool, the cluster assignment matrix can be naturally implemented by the contrast graph as it captures the relatedness of ROIs to the prediction task. Incorporating the contrast graph into graph pooling as group-based prior knowledge can also alleviate overfitting for datasets with limited scales. As shown in Eq. (11), the contrast graph $G_{contrast} = (\boldsymbol{A}_{contrast}, \boldsymbol{X}_{contrast})$ passes through a GNN pooling, $\mathrm{GNN}_{pool}^{(l)}$, to learn a cluster assignment matrix $\boldsymbol{S}^{(l)} \in \mathbb{R}^{m^{(l-1)} \times m^{(l)}}$. Herein, $m^{(l)}$ is a pre-defined number of clusters at layer l controlled by a hyperparameter named the pooling ratio.

$$\boldsymbol{S}^{(l)} = \mathrm{softmax}\left(\mathrm{GNN}_{pool}^{(l)}(\boldsymbol{A}_{contrast}, \boldsymbol{X}_{contrast})\right). \quad (11)$$

To obtain the embedded node feature matrix $\boldsymbol{Z}^{(l)}$ at layer l, we pass the output graph $G^{(l-1)} = (\boldsymbol{A}^{(l-1)}, \boldsymbol{H}^{(l-1)})$ from the previous layer through an embedding GNN:

$$\boldsymbol{Z}^{(l)} = \mathrm{softmax}\left(\mathrm{GNN}_{emb}^{(l)}(\boldsymbol{A}^{(l-1)}, \boldsymbol{H}^{(l-1)})\right), \qquad (12)$$

where the initial representation $\boldsymbol{H}^{(0)}$ is the input feature matrix encoded by a GNN, i.e., $\boldsymbol{H}^{(0)} = \mathrm{GNN}_{enc}(\boldsymbol{X})$.

Once we obtain the cluster assignment matrix $\boldsymbol{S}^{(l)}$ and the embedded node feature matrix $\boldsymbol{Z}^{(l)}$, we generate a coarsened adjacency matrix $\boldsymbol{A}^{(l)}$ and a new feature matrix $\boldsymbol{H}^{(l)}$. This coarsening process can reduce the number of nodes to get higher-level node representations. In particular, we apply the following two equations:

$$\boldsymbol{H}^{(l)} = \boldsymbol{S}^{(l)^\mathsf{T}} \boldsymbol{Z}^{(l)} \in \mathbb{R}^{m^{(l)} \times d}, \qquad (13)$$

$$\boldsymbol{A}^{(l)} = \boldsymbol{S}^{(l)^\mathsf{T}} \boldsymbol{A}^{(l-1)} \boldsymbol{S}^{(l)} \in \mathbb{R}^{m^{(l)} \times m^{(l)}}. \qquad (14)$$

Through Eq. (13 and 14), ContrastPool is able to produce high-quality representations for each input graph by extracting high-level node representations under the guidance of the contrast graph. Intuitively, if an edge with two end ROIs v_1 and v_2 has a high weight in $\boldsymbol{A}_{contrast}$ (i.e., the dual attention of this connection differs significantly in the two groups), the message passing between v_1 and v_2 will be more pronounced. As a result, the scores of v_1 and v_2 in $\boldsymbol{S}$ will share more similarity

and contribute to the set of clusters in a similar manner. With the help of the entropy loss $\mathcal{L}_{E_1}$ (to be introduced in the subsection of "Loss Function"), the learning of the assignment scores of a node will be guided to favor only a few clusters. As a result, two ROIs whose connection is highlighted by the contrast graph will be grouped into the same cluster and high-level node features will be extracted.

Loss Function. Technically, an assignment matrix with entries close to uniform distribution could guide the GNN to treat each ROI and subsequently each node cluster produced in hierarchical pooling equally. Thus, besides the commonly-used cross-entropy loss $\mathcal{L}_{cls}$ [43] for graph classification, we also adopt the loss in [42] to regularize the entropy of the assignment matrix to avoid such equal treatment. Since a fully-connected adjacency matrix of the contrast graph could cause the over-smoothing of GNN, we apply an entropy loss to the adjacency matrix of the contrast graph $\boldsymbol{A}_{contrast}$ to sparsify the matrix. Note that imposing an entropy loss on $\boldsymbol{A}_{contrast}$ is different from thresholding the edges of $\boldsymbol{A}_{contrast}$. The former has an active impact on the learning of $\boldsymbol{A}_{contrast}$ towards the sparse formation, while the latter is a post-process on an already learnt matrix. The two entropy losses shown in Eq. (15 and 16) can help class-indicative ROIs to stand out and subsequently boost the model performance, as evidenced in our ablation study in Section IV-E.

$$\mathcal{L}_{E_1} = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{m^{(l)}} \sum_{i=1}^{m^{(l)}} \text{entropy}\left(\boldsymbol{S}^{(l)}(i,:)\right), \quad (15)$$

$$\mathcal{L}_{E_2} = \frac{1}{m} \sum_{i=1}^{m} \text{entropy}\left(\boldsymbol{A}_{contrast}(i,:)\right). \quad (16)$$

The overall loss $\mathcal{L}$ of ContrastPool is defined as Eq. (17), where λ_1 and λ_2 are trade-off hyperparameters balancing different losses.

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_1 * \mathcal{L}_{E_1} + \lambda_2 * \mathcal{L}_{E_2}. \quad (17)$$

IV. EXPERIMENTAL STUDY

In this section, we first introduce the baseline models and the implementation details of our experiments. We then assess the performance of ContrastPool in comparison with the baseline models. We also present three case studies to show how ContrastPool meets fMRI-specific requirements, and meanwhile to provide the domain interpretation of our dual-attention mechanism. We further conduct ablation studies to analyze the effects of the components in our model. Finally, we perform sensitivity tests on the hyperparameters of our model.

A. Baseline Models

We select various models as baselines, including (1) conventional machine learning models: Logistic Regression, Naïve Bayes, Support Vector Machine Classifier (SVM), k-Nearest Neighbours (kNN) and Random Forest (implemented by the scikit-learn library [44]); (2) general-purposed GNNs: GCN [6], GraphSAGE [45], GIN [40], GAT [18] and GatedGCN [46]; (3) typical graph pooling approaches, DiffPool [42],

SAGPool [25], HGPSLPool [27] and MEWISPool [26]; (4) dense contrast graph with SVM: cs [5]; and (5) neural networks designed for brain networks: BrainNetCNN [4], LiNet [28], PRGNN [12], MG2G [29], BrainGNN [13] and BNTF [34]. For GNN baseline models, we sparsify all input graphs by keeping 20% edges with top correlations in $\boldsymbol{A}$, to avoid over-smoothing.

B. Implementation Details

In ContrastPool, we adopt a GCN layer for GNN_{enc}, a GraphSAGE layer for GNN_{pool} and GNN_{emb}, and a sum pooling with a linear layer for the prediction head. For datasets with more than 2 groups (PPMI and ADNI), we use the most extreme groups to construct the contrast graph: CN and SMC vs. LMCI and AD for the ADNI dataset; NC vs. PD for PPMI. The settings of our experiments mainly follow those in [47]. We split each dataset into 8:1:1 for training, validation and test, respectively. We evaluate each model with the same random seed under 10-fold cross-validation and report the average accuracy. The hyperparameters are grid searched by Table IV.

TABLE IV
HYPERPARAMETER SETTINGS.

batch size	4 (Taowu and Neurocon), 20 (PPMI, ADNI and ABIDE)
λ_1	{10, 1, 0.1}
λ_2	{1e-2, 1e-3, 1e-4}
pooling ratio	{0.3, 0.4, 0.5, 0.6}
L	{2, 3, 4}
learning rate	{0.02, 0.01, 0.005, 0.001}
dropout	{0, 0.1, 0.2}

The whole network is trained in an end-to-end manner using the Adam optimizer [48]. We use the early stopping criterion, i.e., we stop the training once there is no further improvement on the validation loss during 25 epochs. All the codes were implemented using PyTorch [49] and Deep Graph Library [50] packages. All experiments were conducted on a Linux server with an Intel(R) Core(TM) i9-10940X CPU (3.30GHz), a GeForce GTX 3090 GPU, and a 125GB RAM.

C. Comparison with Baselines

We report the accuracy on 5 brain network datasets in Table V. Compared with conventional ML methods, general-purposed GNNs and models for brain network do not show a significant advantage and attain similar performance in most cases. This finding is consistent with the results reported in existing papers [13]. Our ContrastPool outperforms all 21 GNN and ML baselines on all datasets. In particular, ContrastPool improves over all GNNs specifically designed for brain networks by up to 13.6%. The average improvement over GNNs is 8.39%. The p-values of one-sided paired t-tests comparing our ContratPool with the best model of brain networks on two large datasets, ADNI and ABIDE, are 0.0483 and 0.00798, respectively. This indicates that our model significantly outperforms existing methods on these two datasets. However, on small datasets with large standard deviations in 10 folds, t-tests are highly sensitive to outliers existing in the

<p style="text-align:center">TABLE V</p>

GRAPH CLASSIFICATION RESULTS (AVERAGE ACCURACY ± STANDARD DEVIATION) OVER 10-FOLD-CV. THE BEST RESULT IS HIGHLIGHTED IN **BOLD**. THE SECOND BEST RESULT IS <u>UNDERLINED</u>.

	Model	Taowu	PPMI	Neurocon	ADNI	ABIDE
Conventional ML methods	Logistic Regression	**77.50**±7.50	56.50±11.02	68.50±15.17	61.99±0.59	65.82±3.51
	Naïve Bayes	65.00±12.25	58.83±5.42	63.50±11.84	48.27±4.44	63.50±2.69
	SVM	65.00±16.58	<u>63.67</u>±5.11	<u>71.50</u>±13.93	61.77±0.25	60.67±3.61
	kNN	62.50±12.50	53.52±10.34	56.00±21.77	62.59±1.73	60.37±5.64
	Random Forest	57.50±22.50	62.23±4.22	58.50±11.19	61.77±0.25	61.18±5.01
General-purposed GNNs	GCN	60.00±29.15	54.02±9.06	59.00±20.71	61.57±0.60	60.97±2.84
	GraphSAGE	60.00±33.91	55.00±12.89	68.50±15.17	61.19±1.72	63.09±3.11
	GIN	65.00±20.00	57.90±8.12	68.50±15.17	61.87±0.38	57.02±3.88
	GAT	67.50±22.50	54.98±8.03	54.00±15.62	61.34±1.27	60.87±5.02
	GatedGCN	65.00±22.91	52.60±11.51	69.00±25.48	62.06±4.53	63.60±4.70
	DiffPool	65.00±27.84	58.00±11.00	62.50±25.62	<u>63.80</u>±4.64	63.75±3.16
	SAGPool	65.00±25.50	52.14±11.69	65.50±16.95	58.37±3.89	63.70±3.76
	HGPSLPool	60.00±30.00	52.57±11.04	68.50±21.91	59.71±5.79	63.60±3.50
	MEWISPool	57.50±22.50	54.07±12.15	66.00±15.94	61.78±3.10	63.99±4.48
Models for brain networks	cs	**77.50**±17.50	58.36±4.40	63.50±11.84	62.25±0.47	62.59±3.14
	BrainNetCNN	65.00±27.84	57.33±10.32	66.00±22.45	61.08±2.87	<u>65.86</u>±2.36
	LiNet	55.00±24.49	60.71±10.61	56.00±26.91	63.64±1.73	58.14±3.72
	PRGNN	67.50±31.72	58.83±6.89	63.00±23.37	60.71±2.21	60.76±4.12
	MG2G	57.50±22.50	55.45±10.24	68.00±19.77	63.64±5.10	64.41±2.16
	BrainGNN	67.50±25.12	61.71±6.05	56.50±23.03	61.05±1.23	62.88±2.46
	BNTF	65.00±21.08	51.60±6.15	66.00±11.97	61.94±3.11	63.70±4.84
Ours	ContrastPool	**77.50**±17.50	**64.00**±6.63	**75.00**±15.81	**67.08**±2.63	**68.63**±2.65

<p style="text-align:center">TABLE VI</p>

RESULTS OF MORE EVALUATION METRICS ON TAOWU, NEUROCON AND ABIDE DATASETS. THE BEST RESULT IS HIGHLIGHTED IN **BOLD**. FOR MULTICLASS DATASETS OF ADNI AND PPMI, ALL THESE METRICS ARE THE SAME AS ACCURACY IN TABLE V.

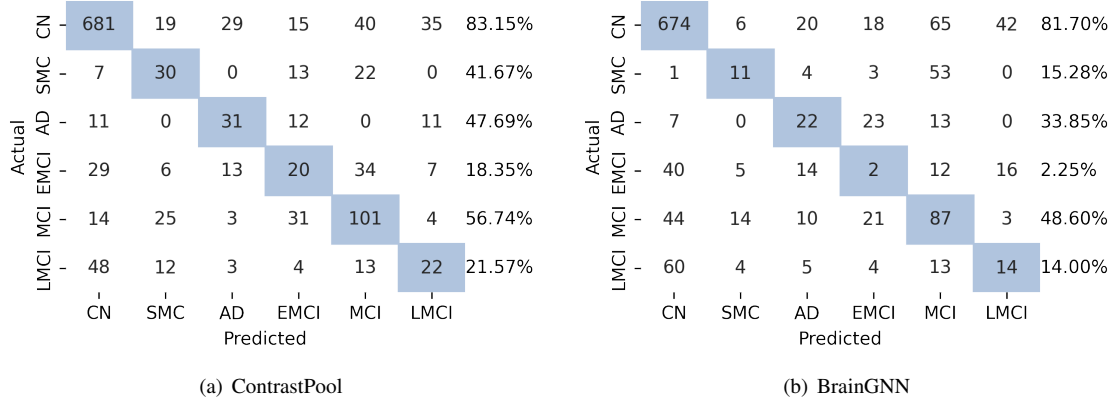|	model	precision	recall	micro-F1	ROC-AUC
Taowu	cs	76.67 ± 25.09	**81.67** ± 14.47	79.09	**77.50** ± 17.50
	ContrastPool	**78.33** ± 23.64	80.00 ± 25.82	**79.16**	**77.50** ± 17.50
Neurocon	SVM	68.33 ± 29.06	85.00 ± 32.02	75.76	65.00 ± 30.00
	ContrastPool	**75.83** ± 18.05	**86.67** ± 20.82	**79.13**	**68.33** ± 20.00
ABIDE	BrainNetCNN	63.79 ± 3.09	64.82 ± 4.90	64.30	66.68 ± 2.53
	ContrastPool	**64.48** ± 4.08	**66.10** ± 9.13	**65.28**	**68.16** ± 3.61

(a) ContrastPool (b) BrainGNN

Fig. 2. Confusion matrices with class-wise accuracy of ContrastPool and BrainGNN on the ADNI dataset. Each confusion matrix is obtained by adding the confusion matrices on test sets of all folds.

10-fold results, thus decreasing the t-statistic calculated and lowering the chance of rejecting the null hypothesis.

It is observed from Table V that different methods exhibit varied standard deviations on different datasets. Neural network models typically have a larger number of parameters than conventional ML methods, making them susceptible to overfitting, especially on small datasets like Taowu. Consequently, these methods typically exhibit higher standard deviations than conventional ML methods when tested on the same dataset. Regarding the comparison of standard deviation across different datasets, large datasets such as ABIDE are more likely to guarantee consistent data distribution across different folds, and thus tend to achieve smaller standard deviations than small datasets such as Taowu and Neurocon. Remarkably, the standard deviation obtained by our ContrastPool lies at the lower end among all with general GNNs and models for brain networks on all datasets. This also demonstrates that ContrastPool is able to remit the problem of overfitting.

(a) ContrastPool

(b) BrainGNN

Fig. 3. Confusion matrices with class-wise accuracy of ContrastPool and BrainGNN on the PPMI dataset.

(a) On the ABIDE dataset. Blue edges denote higher attention on TC group and red edges denote higher attention on ASD group.

(b) On the ADNI dataset. Blue edges denote higher attention on CN/SMC group and red edges denote higher attention on AD/LMCI group.

(c) On the PPMI dataset. Blue edges denote higher attention on NC group and red edges denote higher attention on PD group.
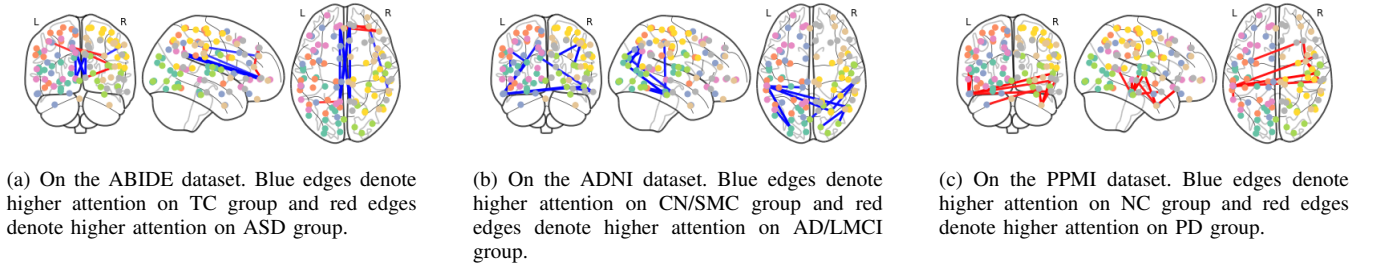
Fig. 4. Contrast graph visualization. The connections with the top 10 ROI-wise attention weights are highlighted.

Apart from accuracy, we also report other evaluation metrics, including precision, recall, micro-F1, and ROC-AUC, of the top 2 models on each dataset. As shown in Table VI, ContrastPool performs the best on all datasets over all these metrics except for a single case (recall on Taowu when compared with cs). Note that we do not report the additional metrics on the two multi-class datasets PPMI and ADNI. This is because in the multi-class case, all these metrics are the same as accuracy, with the superiority of ContrastPool over all baselines already demonstrated in Table V.

To further analyze the performance on multi-class datasets, we plot the confusion matrices with class-wise accuracy on ADNI and PPMI datasets for ContrastPool and BrainGNN in Figs. 2 and 3. We can observe that although both ContrastPool and BrainGNN tend to predict subjects as the majority class (CN on ADNI and PD on PPMI), our ContrastPool exhibits higher accuracy on the minority classes (see the column beside each matrix).

These experimental results demonstrate the effectiveness of our brain network oriented model design. Existing GNNs with attention mechanism, such as GAT, PRGNN, and BrainGNN, only utilize attention over ROI-level of each subject's feature matrix. To the best of our knowledge, our ContrastPool is the first method that applies ROI- and subject-wise attention on both adjacency matrices and feature matrices.

D. Case Studies

In this subsection, we present case studies for ROI-wise attention, subject-wise attention, assignment matrix and generalization performance to showcase how ContrastPool meets the need of three characteristics of fMRI data.

Interpretation of ROI-wise Attention. To interpret the rationality of ContrastPool, we visualize the learnt contrast graphs on ABIDE, ADNI, and PPMI datasets. In order to visualize the ROI-wise attentions that are discriminative across different groups, we plot out the ROI-wise attention scores stored in the contrast graph. Essentially, each ROI attention in the contrast graph is obtained by averaging over all subjects (Eq. (10)) and contrasting between two groups (Eq. (4)). As shown in Fig. 4, we select the edges with the top 10 ROI-wise attention weights. The chord diagram in Fig. 5 displays the edges with top-20 ROI-wise attention scores. We merge the ROIs in the same area in the chord diagram for a clearer visualization. Take the ABIDE dataset in Figs. 4(a) and 5(a) as an example. We can see that the contrast graph differentiates ROI pairs (edges in brain networks) with various importance levels (attention weights) and highlights a number of functional connections with high importance in distinguishing subjects from ASD and TC. Connections between the prefrontal cortex, parietal and cingulate are highlighted by our ROI-wise attention. Some ASD-specific neural machianism [51] may be underlying these connections, and these ROIs were regarded as key involved regions in previous ASD studies [52], [53]. Similar ROI-wise interpretations are found on Alzheimer's and Parkinson's as well. ROIs related to parietal and posterior are highlighted on ADNI (as shown in Figs. 4(b) and 5(b)), while those within temporal and ventral prefrontal cortex are highlighted on PPMI (as shown in Figs. 4(c) and 5(c)). These findings also match the domain knowledge in prior research of Alzheimer's [54]–[57] and Parkinson's [58]–[60].

We also find some highlighted ROIs that diverge from conventional neuroscientific understanding. For example, pos-

(a) On the ABIDE dataset.　　(b) On the ADNI dataset.　　(c) On the PPMI dataset.
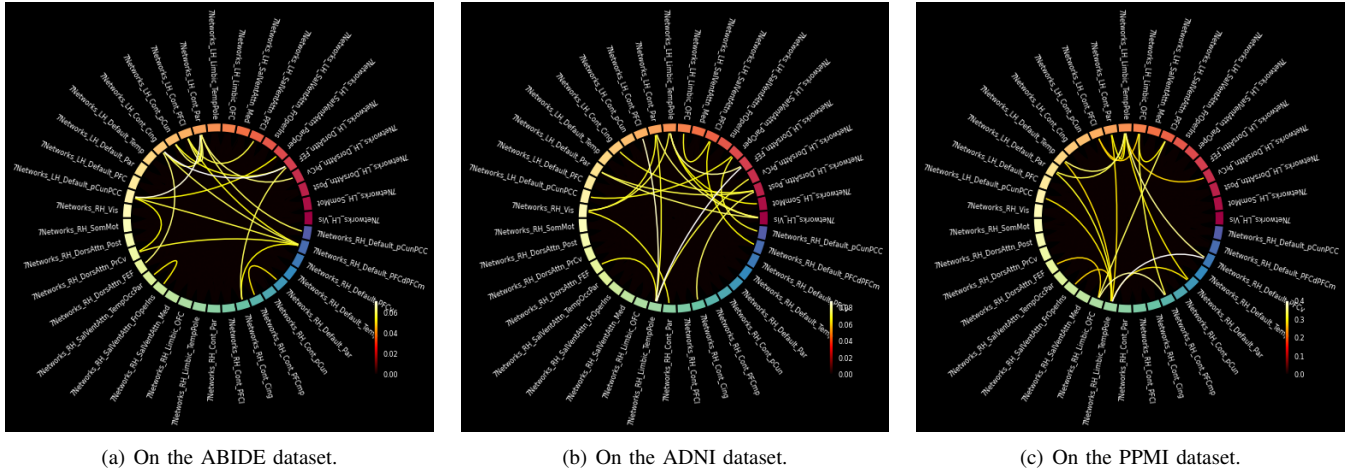
Fig. 5. Chord diagrams of contrast graphs. Only the edges with top-20 ROI-wise attention scores are shown for better visualization. (a) ROIs related to prefrontal cortex, parietal and cingulate are highlighted for Autism. (b) ROIs related to parietal and posterior are highlighted for Alzheimer's. (c) ROIs related to temporal and ventral prefrontal cortex are highlighted for Parkinson's.

terior has a high attention weight in our model on the ADNI dataset. This area belongs to the dorsal attention network, which may imply Alzheimer's Disease is related to the function of goal-directed, voluntary control of visuospatial attention [61]. This insight has not been identified by existing literature [62]. The attention weights highlighted by Contrast-Pool may shed some light on neurological pathologies from a data-driven perspective and facilitate related clinical research with unconventional hints.

(a) **Site distribution on original features.**　　(b) **Site distribution on the representations of ContrastPool w/o subject-wise attention.**
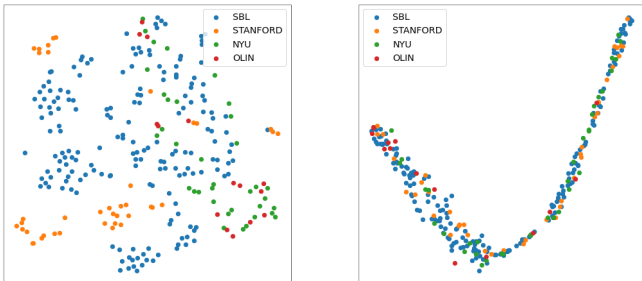
Fig. 6. The visualization of the original features and ContrastPool representations for subjects from multiple sites on the ABIDE dataset. Each point in the figure represents a subject and different colors denote the sites these subjects are acquired from. The representation of each subject is obtained by mean pooling and visualized by t-SNE [63]. Compared with (b), (a) exhibits obvious site distribution shifts.

We further conducted an experiment to verify the power of the ROI-wise attention in reducing the site-specific noise. We used the ABIDE dataset, which consists of subjects collected from multiple sites and thus may exhibit site differences (scanner variability, different inclusion/exclusion criteria). Fig. 6 in this document provides a visualization of the site distribution of subjects in the ABIDE dataset under (a) original features and (b) the representations generated by ContrastPool with subject-wise attention disabled. Each point in the figure represents a subject, and different colors denote the sites from

which these subjects are acquired. Fig. 6(a) shows that subjects from different sites do not span over the same subspace but different subspaces. This implies that the site-specific noise may contribute to the distribution shifts under the original features. On the contrary, when we plot out the site distribution under the representations obtained by ContrastPool in Fig. 6(b), we observe that such distribution shifts are alleviated: the subjects from different sites span over the same subspace in a more uniform manner. We disable the subject-wise attention so as to focus on the impact from the ROI-wise attention (Eq. (8)). We remark that site information is not utilized at all in the model training and yet ContrastPool is able to reduce such site-specific noise via the design of the ROI-wise attention.

Subject-wise Attention. We design an experiment on the ADNI dataset to demonstrate ContrastPool's capability of underscoring informative subjects. Specifically, we merge the MCI group (82 subjects) with the AD group (143 subjects) into a single group (without revealing to ContrastPool the exact group each subject is from) to be used to contrast against the CN group. Specifically, for a better visualization of the subject-wise attention, we aggregate the weights of all ROIs for each subject into a scalar to represent its subject-wise attention level. Fig. 7(a) shows the subjects with the top 50 subject-wise attentions and their class distribution. The proportion of AD subjects in the top 50 attention subjects (35/50) is much higher than its proportion in the merged group of subjects (143/225). The top 5 highest attention subjects are all from the AD group. This observation illustrates that subject-wise attention could lead the model to focus more on typical/representative subjects in the dataset.

A similar conclusion can be drawn on PPMI (the other multi-class dataset) as well. We merge the SWEDD group (12 subjects) with the PD group (89 subjects) into a single group to be used to contrast against the NC group. Fig. 7(b) shows the subjects with the top 30 attention weights. The proportion of PD subjects in the top 30 attention subjects (28/30) is much higher than its proportion in the merged group of subjects (89/101). The top 10 highest attention subjects are all from the
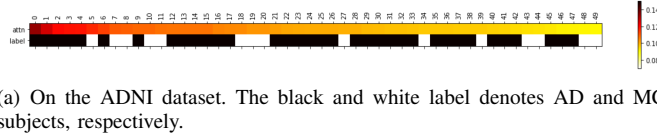
(a) On the ADNI dataset. The black and white label denotes AD and MCI subjects, respectively.
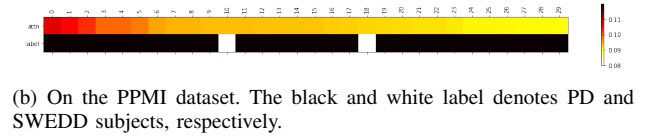
(b) On the PPMI dataset. The black and white label denotes PD and SWEDD subjects, respectively.

Fig. 7. The heatmap and class distribution on top 50 and top 30 subjects in the subject-wise attention on ADNI and PPMI datasets, respectively.

PD group, which once again demonstrates the ability of ContrastPool to automatically highlight representative subjects.

Assignment Matrix. To better understand how the contrast graph works in the differentiable graph pooling, we provide a visualization of the assignment matrix $S^{(1)}$ at the first layer of ContrastPool on the PPMI dataset in Fig. 8. Each row corresponds to an ROI and each column is a cluster. We observe that the differentiable pooling is able to emphasize the most important ROIs and enforce the scores of most of the other ROIs to be close to 0. The entries with high values belong to temporal and ventral prefrontal cortex, which are consistent with the ROIs that have been highlighted in the contrast graph. This demonstrates that the differentiable pooling is well guided by the contrast graph in highlighting the discriminative ROIs.
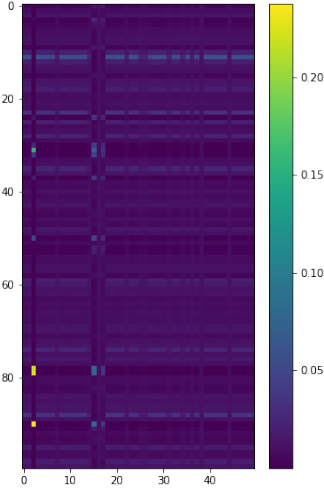
Fig. 8. Visualization of the assignment matrix $S^{(1)}$ at the first layer of ContrastPool on PPMI.

Generalization Performance. We also observe that ContrastPool is able to alleviate the overfitting problem, which is a common problem when applying GNNs to brain networks. An example is shown in Fig. 9, where we plot the accuracy curves of ContrastPool and DiffPool on the ABIDE dataset. The test accuracy of DiffPool decreases after its training accuracy coverages to 1. ContrastPool dramatically narrows the gap between the training and test accuracy, which demonstrates its ability in remitting the overfitting problem.

E. Ablation Studies

In this subsection, we validate empirically the design choices made in different components of our model: 1) the dual-attention mechanism and the contrast operation in the Contrastive Dual-Attention (CDA) block; and (2) the
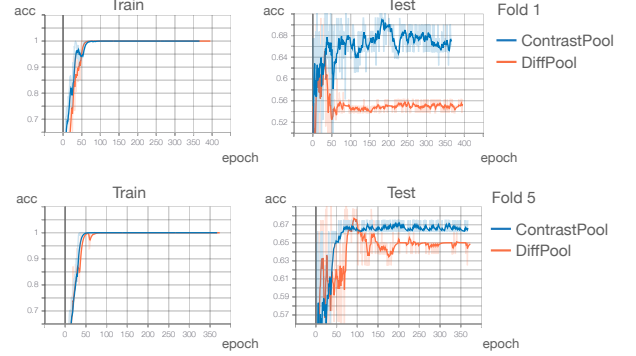
Fig. 9. Accuracy curves of ContrastPool and DiffPool on two folds on the ABIDE dataset.

loss function. All experiments are conducted on the ABIDE dataset.

Contrastive Dual-attention Block. Our CDA block performs dual attention (subject-wise and ROI-wise) on both adjacency matrices and feature matrices. We disable each of them to inspect their effects. The results reported in Table VII demonstrate that the CDA with all attentions enabled achieves the best performance. In particular, if the dual attention is entirely disabled (row 1), the summary graph is essentially obtained by averaging over all graphs in each group, which compromises the performance. Moreover, relying solely on ROI-wise attention (row 2) may lead to a further deterioration in performance. This implies that without the support of subject-wise attention, the ROI-wise attention suffers from the information extracted from less representative subjects.

To verify the effectiveness of the contrast operation, we conduct an experiment with a CDA variant without considering groups. The results (the last two rows of Table VII) show that ContrastPool with the contrast operation obtains significantly better performance. This verifies the necessity of capturing the discrepancy across groups in the model design.

TABLE VII
ABLATION STUDY ON CDA BLOCK ON ABIDE. WINNER IS
HIGHLIGHTED IN **BOLD**.

Adjacency Matrices		Feature Matrices		Contrast	acc ± std
subject-wise	ROI-wise	subject-wise	ROI-wise		
				✓	65.88 ± 4.11
	✓		✓	✓	64.63 ± 3.83
✓		✓		✓	66.25 ± 2.68
		✓	✓	✓	66.88 ± 4.23
✓	✓			✓	67.88 ± 4.78
✓	✓	✓	✓		62.71 ± 3.08
✓	✓	✓	✓	✓	**68.63 ± 2.65**

Loss Function. We test our design of the loss function by disabling two entropy losses. As shown in Table VIII, the

results demonstrate that both $\mathcal{L}_{E_1}$ and $\mathcal{L}_{E_2}$ are effective in boosting the model performance.

TABLE VIII
ABLATION STUDY ON ENTROPY LOSS ON ABIDE. THE BEST RESULT IS HIGHLIGHTED IN **BOLD**.

$\mathcal{L}_{cls}$	$\mathcal{L}_{E_1}$	$\mathcal{L}_{E_2}$	acc ± std
✓			64.88 ± 4.45
✓	✓		66.13 ± 2.65
✓	✓	✓	**68.63** ± 2.65

F. Hyperparameter Analysis

In this subsection, we study the sensitivity of three important hyperparameters in ContrastPool, which are the trade-off parameters in Eq. (17), the pooling ratio and the number of layers. All experiments are conducted on the ABIDE dataset. **Trade-off Parameters λ_1 and λ_2.** These two hyperparameters are used in the overall loss function $\mathcal{L}$ (Eq. 17) for the trade-off between the classification loss and the two entropy losses. We tune the value of λ_1 from 10 to 0.1 and λ_2 from 1e-2 to 1e-4. The results presented in Fig. 10 show that our model performs the best when $\lambda_1 = 1$ and $\lambda_2 = 1e-3$. The same optimal values of λ_1 and λ_2 are found on other datasets.
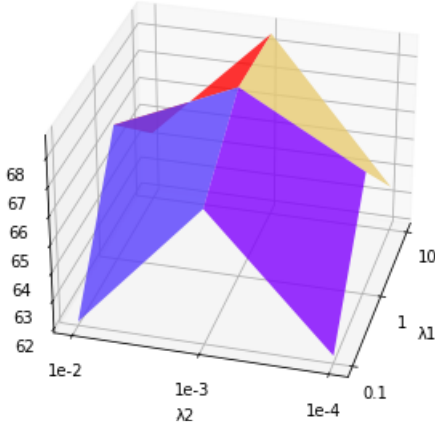
Fig. 10. Results when tuning **λ_1** and **λ_2** on ABIDE.

Pooling Ratio. The number of nodes $m^{(l)}$ in the output graph of each ContrastPool layer is controlled by the pooling ratio. A smaller pooling ratio would lead to fewer node clusters in each ContrastPool layer. Herein, we tune it from 0.3 to 0.6. As shown in Fig. 11, the best choice of pooling ratio is 0.4. The best pooling ratio on different datasets varies slightly within the range of [0.4, 0.5].
Number of Layers. The depth of the neural network can undoubtedly affect the model performance. We vary the number of layers L in ContrastPool from 2 to 4, and report the results in Fig. 11. ContrastPool achieves the best performance when we set L to 2. This indicates that by leveraging the contrast graph, our ContrastPool requires fewer layers to obtain good representations, while most other GNN baselines need to be deeper (e.g., 4 layers) to achieve best performance. The same conclusion of the optimal L can be drawn on other datasets.
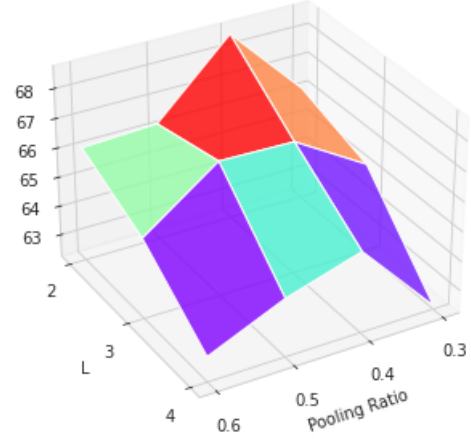
Fig. 11. Results when tuning pooling ratio and the number of layers on ABIDE.

V. CONCLUSION

This paper proposes a novel GNN-based solution for brain network classification, taking the unique characteristics of the underlying fMRI data into account. Our proposed method, *ContrastPool*, can adaptively select the most discriminative regions of interest and the most representative subjects by engaging a contrastive dual-attention block. It allows for a flexible local information aggregation within each group. We demonstrate the superiority of our method over 21 state-of-the-art baselines on 5 brain-network datasets spanning over 3 diseases. Moreover, our case studies show the interestingness, simplicity, and high explainability of the patterns extracted by our method, which find consistency in the neuroscience literature. We hope our work can inspire further research in leveraging GNNs for brain network analysis, and show significance in real-world applications, such as the early diagnosis and personalized treatments of neurodegenerative diseases.

REFERENCES

[1] K. J. Worsley, C. H. Liao, J. Aston, V. Petre, G. Duncan, F. Morales, and A. C. Evans, "A general statistical analysis for fmri data," *Neuroimage*, vol. 15, no. 1, pp. 1–15, 2002.

[2] R. A. Poldrack, Y. O. Halchenko, and S. J. Hanson, "Decoding the large-scale structure of brain function by classifying mental states across individuals," *Psychological science*, vol. 20, no. 11, pp. 1364–1372, 2009.

[3] J. Xu, Y. Yang, D. T. J. Huang, S. S. Gururajapathy, Y. Ke, M. Qiao, A. Wang, H. Kumar, J. McGeown, and E. Kwon, "Data-driven network neuroscience: On data collection and benchmark," in *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023.

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

[5] T. Lanciano, F. Bonchi, and A. Gionis, "Explainable classification of brain networks via contrast subgraphs," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3308–3318.

[6] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[7] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *International conference on machine learning*. PMLR, 2017, pp. 1263–1272.

[8] J. Xu, A. Zhang, Q. Bian, V. P. Dwivedi, and Y. Ke, "Union subgraph neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 14, 2024, pp. 16 173–16 183.

[9] S. Wu, F. Sun, W. Zhang, X. Xie, and B. Cui, "Graph neural networks in recommender systems: a survey," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–37, 2022.

[10] Q. Bian, J. Xu, H. Fang, and Y. Ke, "Cpmr: Context-aware incremental sequential recommendation with pseudo-multi-task learning," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 120–130.

[11] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Distance metric learning using graph convolutional networks: Application to functional brain networks," in *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*. Springer, 2017, pp. 469–477.

[12] X. Li, Y. Zhou, N. C. Dvornek, M. Zhang, J. Zhuang, P. Ventola, and J. S. Duncan, "Pooling regularized graph neural network for fmri biomarker analysis," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23*. Springer, 2020, pp. 625–635.

[13] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L. H. Staib, P. Ventola, and J. S. Duncan, "Braingnn: Interpretable brain graph neural network for fmri analysis," *Medical Image Analysis*, vol. 74, p. 102233, 2021.

[14] Y. Yan, J. Zhu, M. Duda, E. Solarz, C. Sripada, and D. Koutra, "Groupinn: Grouping-based interpretable neural network for classification of limited, noisy brain data," in *proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 772–782.

[15] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," *Advances in neural information processing systems*, vol. 33, pp. 5812–5823, 2020.

[16] Y. Yin, Q. Wang, S. Huang, H. Xiong, and X. Zhang, "Autogcl: Automated graph contrastive learning via learnable view generators," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 8, 2022, pp. 8892–8900.

[17] J. Chen and G. Kou, "Attribute and structure preserving graph contrastive learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 6, 2023, pp. 7024–7032.

[18] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[19] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" *arXiv preprint arXiv:2105.14491*, 2021.

[20] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun, "Gram: graph-based attention model for healthcare representation learning," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 787–795.

[21] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao, "Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 1903–1911.

[22] K. K. Thekumparampil, C. Wang, S. Oh, and L.-J. Li, "Attention-based graph neural network for semi-supervised learning," *arXiv preprint arXiv:1803.03735*, 2018.

[23] J. B. Lee, R. Rossi, and X. Kong, "Graph classification using structural attention," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 1666–1674.

[24] Q. Sun, J. Li, H. Peng, J. Wu, Y. Ning, P. S. Yu, and L. He, "Sugar: Subgraph neural network with reinforcement pooling and self-supervised mutual information mechanism," in *Proceedings of the Web Conference 2021*, 2021, pp. 2081–2091.

[25] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *International conference on machine learning*. PMLR, 2019, pp. 3734–3743.

[26] A. Nouranizadeh, M. Matinkia, M. Rahmati, and R. Safabakhsh, "Maximum entropy weighted independent set pooling for graph neural networks," *arXiv preprint arXiv:2107.01410*, 2021.

[27] Z. Zhang, J. Bu, M. Ester, J. Zhang, C. Yao, Z. Yu, and C. Wang, "Hierarchical graph pooling with structure learning," *arXiv preprint arXiv:1911.05954*, 2019.

[28] X. Li, N. C. Dvornek, Y. Zhou, J. Zhuang, P. Ventola, and J. S. Duncan, "Graph neural network for interpreting task-fmri biomarkers," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22*. Springer, 2019, pp. 485–493.

[29] M. Xu, D. L. Sanz, P. Garces, F. Maestu, Q. Li, and D. Pantazis, "A graph gaussian embedding method for predicting alzheimer's disease progression with meg brain networks," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 5, pp. 1579–1588, 2021.

[30] H. Zhang, R. Song, L. Wang, L. Zhang, D. Wang, C. Wang, and W. Zhang, "Classification of brain disorders in rs-fmri via local-to-global graph neural networks," *IEEE Transactions on Medical Imaging*, 2022.

[31] J.-E. Ding, C.-C. Hsu, and F. Liu, "Parkinson disease classification using contrastive graph cross-view learning with multimodal fusion of spect images and clinical features," *arXiv preprint arXiv:2311.14902*, 2023.

[32] B.-H. Kim, J. C. Ye, and J.-J. Kim, "Learning dynamic graph representation of brain connectome with spatio-temporal attention," *Advances in Neural Information Processing Systems*, vol. 34, pp. 4314–4327, 2021.

[33] Y. Yu, X. Kan, H. Cui, R. Xu, Y. Zheng, X. Song, Y. Zhu, K. Zhang, R. Nabi, Y. Guo *et al.*, "Learning task-aware effective brain connectivity for fmri analysis with graph neural networks," *arXiv preprint arXiv:2211.00261*, 2022.

[34] X. Kan, W. Dai, H. Cui, Z. Zhang, Y. Guo, and C. Yang, "Brain network transformer," *arXiv preprint arXiv:2210.06681*, 2022.

[35] L. Badea, M. Onu, T. Wu, A. Roceanu, and O. Bajenaru, "Exploring the reproducibility of functional connectivity alterations in parkinson's disease," *PLoS One*, vol. 12, no. 11, p. e0188196, 2017.

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

[37] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham *et al.*, "The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives," *Frontiers in Neuroinformatics*, vol. 7, p. 27, 2013.

[38] O. Esteban, C. J. Markiewicz, R. W. Blair, C. A. Moodie, A. I. Isik, A. Erramuzpe, J. D. Kent, M. Goncalves, E. DuPre, M. Snyder *et al.*, "fmriprep: a robust preprocessing pipeline for functional mri," *Nature methods*, vol. 16, no. 1, pp. 111–116, 2019.

[39] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Yeo, "Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri," *Cerebral cortex*, vol. 28, no. 9, pp. 3095–3114, 2018.

[40] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

[41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[42] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," *Advances in neural information processing systems*, vol. 31, 2018.

[43] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 20, no. 2, pp. 215–232, 1958.

[44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[45] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

[46] X. Bresson and T. Laurent, "Residual gated graph convnets," *arXiv preprint arXiv:1711.07553*, 2017.

[47] V. P. Dwivedi, C. K. Joshi, A. T. Luu, T. Laurent, Y. Bengio, and X. Bresson, "Benchmarking graph neural networks," *arXiv preprint arXiv:2003.00982*, 2020.

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[49] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

[50] M. Wang, L. Yu, D. Zheng, Q. Gan, Y. Gai, Z. Ye, M. Li, J. Zhou, Q. Huang, C. Ma *et al.*, "Deep graph library: Towards efficient and scalable deep learning on graphs." 2019.

[51] S.-J. Weng, J. L. Wiggins, S. J. Peltier, M. Carrasco, S. Risi, C. Lord, and C. S. Monk, "Alterations of resting state functional connectivity in the default network in adolescents with autism spectrum disorders," *Brain research*, vol. 1313, pp. 202–214, 2010.

[52] M. Assaf, K. Jagannathan, V. D. Calhoun, L. Miller, M. C. Stevens, R. Sahl, J. G. O'Boyle, R. T. Schultz, and G. D. Pearlson, "Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients," *Neuroimage*, vol. 53, no. 1, pp. 247–256, 2010.

[53] R. K. Kana, E. B. Sartin, C. Stevens Jr, H. D. Deshpande, C. Klein, M. R. Klinger, and L. G. Klinger, "Neural networks underlying language and social cognition during self-other processing in autism spectrum disorders," *Neuropsychologia*, vol. 102, pp. 116–123, 2017.

[54] R. L. Gould, R. G. Brown, A. M. Owen, E. T. Bullmore, and R. J. Howard, "Task-induced deactivations during successful paired associates learning: an effect of age but not alzheimer's disease," *Neuroimage*, vol. 31, no. 2, pp. 818–831, 2006.

[55] M. R. Brier, J. B. Thomas, A. Z. Snyder, T. L. Benzinger, D. Zhang, M. E. Raichle, D. M. Holtzman, J. C. Morris, and B. M. Ances, "Loss of intranetwork and internetwork resting state functional connections with alzheimer's disease progression," *Journal of Neuroscience*, vol. 32, no. 26, pp. 8890–8899, 2012.

[56] P. Alexopoulos, C. Sorg, A. Förschler, T. Grimmer, M. Skokou, A. Wohlschläger, R. Perneczky, C. Zimmer, A. Kurz, and C. Preibisch, "Perfusion abnormalities in mild cognitive impairment and mild dementia in alzheimer's disease measured by pulsed arterial spin labeling mri," *European archives of psychiatry and clinical neuroscience*, vol. 262, pp. 69–77, 2012.

[57] S. Tu, S. Wong, J. R. Hodges, M. Irish, O. Piguet, and M. Hornberger, "Lost in spatial translation–a novel tool to objectively assess spatial disorientation in alzheimer's disease and frontotemporal dementia," *Cortex*, vol. 67, pp. 83–94, 2015.

[58] O. Monchi, M. Petrides, J. Doyon, R. B. Postuma, K. Worsley, and A. Dagher, "Neural bases of set-shifting deficits in parkinson's disease," *Journal of Neuroscience*, vol. 24, no. 3, pp. 702–710, 2004.

[59] N. J. Gerrits, Y. D. van der Werf, K. M. Verhoef, D. J. Veltman, H. J. Groenewegen, H. W. Berendse, and O. A. van den Heuvel, "Compensatory fronto-parietal hyperactivation during set-shifting in unmedicated patients with parkinson's disease," *Neuropsychologia*, vol. 68, pp. 107–116, 2015.

[60] M. A. Fernández-Seara, E. Mengual, M. Vidorreta, G. Castellanos, J. Irigoyen, E. Erro, and M. A. Pastor, "Resting state functional connectivity of the subthalamic nucleus in p arkinson's disease assessed using arterial spin-labeled perfusion f mri," *Human brain mapping*, vol. 36, no. 5, pp. 1937–1950, 2015.

[61] M. D. Fox, M. Corbetta, A. Z. Snyder, J. L. Vincent, and M. E. Raichle, "Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems," *Proceedings of the National Academy of Sciences*, vol. 103, no. 26, pp. 10 046–10 051, 2006.

[62] M. F. Mendez, M. Ghajarania, and K. M. Perryman, "Posterior cortical atrophy: clinical characteristics and differences compared to alzheimer's disease," *Dementia and geriatric cognitive disorders*, vol. 14, no. 1, pp. 33–40, 2002.

[63] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.