BrainMass: Advancing Brain Network Analysis for Diagnosis with Large-scale Self-Supervised Learning

Yanwu Yang, *Member, IEEE*, Chenfei Ye, Guinan Su, Ziyao Zhang, Zhikai Chang, Hairui Chen, Piu Chan, Yue Yu, Ting Ma *Member, IEEE*

***Abstract*— Foundation models pretrained on large-scale datasets via self-supervised learning demonstrate exceptional versatility across various tasks. Due to the heterogeneity and hard-to-collect medical data, this approach is especially beneficial for medical image analysis and neuroscience research, as it streamlines broad downstream tasks without the need for numerous costly annotations. However, there has been limited investigation into brain network foundation models, limiting their adaptability and generalizability for broad neuroscience studies. In this study, we aim to bridge this gap. In particular, (1) we curated a comprehensive dataset by collating images from 30 datasets, which comprises 70,781 samples of 46,686 participants. Moreover, we introduce pseudo-functional connectivity (pFC) to further generates millions of augmented brain networks by randomly dropping certain timepoints of the BOLD signal. (2) We propose the BrainMass framework for brain network self-supervised learning via mask modeling and feature alignment. BrainMass employs Mask-ROI Modeling (MRM) to bolster intra-network dependencies and regional specificity. Furthermore, Latent Representation Alignment (LRA) module is utilized to regularize augmented brain networks of the same participant with similar topological properties to yield similar latent representations by aligning their latent embeddings. Extensive experiments on eight internal tasks and seven external brain disorder diagnosis tasks show BrainMass's superior performance, highlighting its significant generalizability and adaptability. Nonetheless, BrainMass demonstrates powerful few/zero-shot learning abilities and exhibits meaningful interpretation to various diseases, showcasing its potential use for clinical applications.**

***Index Terms*— Self-supervised learning, brain network, Transformer, large-scale, pretrain**

This work was done by Yanwu Yang during his internship at Peng Cheng Laboratory. The study is supported by grants from the National Natural Science Foundation of China (62276081), The National Key Research and Development Program of China (2021YFC2501202), and the Major Key Project of PCL. Corresponding author: Yue Yu, Ting Ma.

Yanwu Yang and Chenfei Ye contributed equally to this work.

Yanwu Yang, and Hairui Chen are with the School of Electronics and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, China, and the Peng Cheng Laboratory, Shenzhen, Guangdong, China. (e-mail: 20b952019@stu.hit.edu.cn, 23b952017@stu.hit.edu.cn)

Guinan Su is with the Tencent Data Platform, Shenzhen 518057, China. (E-mail: guinansu33@gmail.com)

Ting Ma is with the School of Electronics and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, China, the Peng Cheng Laboratory, Shenzhen, Guangdong, China, and International Research Institute for Artificial Intelligence, Harbin Institute of Technology (Shenzhen), Shenzhen, China. (e-mail: tma@hit.edu.cn)

Ziyao Zhang is with the Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong 518000, China and Peng Cheng Laboratory, Shenzhen, Guangdong 518000, China. (e-mail: zhangzy@pcl.ac.cn)

Yue Yu is with the Peng Cheng Laboratory, Shenzhen, Guangdong, China. (e-mail: yuy@pcl.ac.cn)

Piu Chan is with Xuanwu Hospital, Capital Medical University, Beijing, China. (E-mail: pbchan@hotmail.com)

Chenfei Ye and Zhikai Chang are with the Harbin Institute of Technology at Shenzhen, Shenzhen, China (e-mail: Chenfei.ye@foxmail.com, 22b352010@stu.hit.edu.cn)

I. Introduction

Functional Magnetic Resonance Imaging (fMRI), utilizing the blood-oxygen-level-dependent (BOLD) effect, has become an instrumental tool in neuroscience. It offers a unique opportunity to map the neural substrates of cognition in vivo [1]–[3]. Recently, fMRI has been widely used to analyze brain dysfunctions and can reveal networks of interacting brain regions. Many brain disorders appear to originate from disruptions confined to specific brain functions, rather than from structural focal lesions [4]–[6]. A key outcome of this paradigm is the development of functional brain networks, which are established through correlations between BOLD signal from various regions of interest (ROIs) to estimate the neural interactions and temporal synchronization. These networks have become indispensable tools for brain disorder analysis, examining the underlying disconnectome in various diseases [7], [8].

In recent years, the field of brain functional network analysis has been greatly influenced by deep learning approaches, which characterize complex interactions of ROIs with nonlinear and deep embedded representations, and significantly improves the disease diagnosis performance. These include a range of techniques such as convolutional neural networks (CNN) [9]–[11], graph neural networks (GNN) [12]–[15], and Transformer networks [16], [17]. Despite significant progress, a pervasive limitation of these studies is their limited generalizability and adaptability [18], [19]. Task-specific models are still the main methods used, that are limited by the number of annotated samples and poor adaptation to other tasks. And the lack of capabilities for few-shot or zero-shot learning, limits their potential use in clinical scenarios where only a

few MRIs with annotations are available. Moreover, the data heterogeneity also hampers the generalizability [20], [21].

One way to address this issue is through large-scale self-supervised learning (SSL) to produce homogeneous and generalizable representations. This method has shown promise, leading to impressive performance gains in a wide variety of downstream tasks across other domains [22]–[25]. Unlike traditional pretrained models, foundation models pre-trained on large-scale datasets can handle a wide variety of tasks with a single set of model weights [26]. However, in the field of medical image analysis, developing foundation models, in particular for brain networks, presents a significant challenge due to the limited data samples and insufficient self-supervised learning. Current studies leveraging SSL for brain network only achieve comparable performance to non-SSL methods [18], [19], [27]. Consequently, specific foundation models on brain network is urgently needed in this field at the moment.

To this end, we aim to bridge the gap in foundation models for brain networks. In this paper, we curated a large cohort comprising 70,781 samples of 46,686 participants in multiple centers. We also introduce an augmentation method to create more brain networks that involves randomly dropping time-points in the BOLD signals to pseudo-functional connectivity (pFC). Moreover, we propose BrainMass, the first foundation model specifically designed for **Brain** network analysis with **M**ask modeling and representation **A**lignment via **S**elf-**S**upervised learning to pre-train the Transformer encoder:

(1) **MRM**: MRM is executed by randomly masking some ROIs and predicting the masked features by the remaining. In particular, classification heads are utilized to predict the meta labels (indices of the masked ROIs), and reconstruction heads are deployed to estimate the features of the masked ROIs. This inclusion helps relate intra-network dependencies and enhances locality characteristics for downstream tasks.

(2) **LRA**: BrainMass leveraging LRA employs a dual-branch approach to extract representations from two pFCs derived from the same BOLD signal and regularizes them to achieve similar latent embeddings. This design acknowledges that augmented brain networks derived from the same participant should yield similar latent representations. We leverage a dual branch network to extract the embeddings of two pFCs and regularize the them to be closer.

To evaluate the effectiveness of our BrainMass, eight internal and seven external diagnosis tasks were carried out as the downstream tasks. We extracted the learned representations from BrainMass and employed an SVM classifier for classification. Our extensive experimental results demonstrated that BrainMass not only outperformed existing models but also exhibited remarkable generalizability, adaptability, and few-shot capabilities. Our key contributions are outlined as follows:

- We built a large cohort comprising 46,686 participants with 70,781 samples for large-scale brain network studies. Furthermore, we developed augmented brain networks using pseudo-functional connectivity (pFC) to further enlarge the training set.
- We introduced the first brain network foundation model, BrainMass, and demonstrated its superior diagnostic performance, along with its impressive generalizability and adaptability across eight internal and seven external tasks.
- Our explanatory analysis revealed that BrainMass is capable of identifying the patterns of various brain disorders and pinpointing meaningful key biomarkers.
- Our BrainMass exhibits powerful capabilities in zero-shot and few-shot learning, showcasing its potential for clinical applications.

Our project is publicly online at `https://github.com/podismine/BrainMass`. The pre-trained weights are available for researchers to easily adapt the model for various tasks and analyze the biomarkers without the need for computationally expensive supervised fine-tuning.

II. Related Works

A. Brain network study

Significant advancements have been made over the past decade in the application of neuroimaging techniques to uncover alterations in brain network associated with various brain disorders [28]–[30]. Convolutional neural networks (CNN) are firstly proposed to facilitate end-to-end disease identification with promising performances and have been widely applied for analyzing network patterns such as BrainNetCNN [9] and Deep Convolutional Auto-Encoder [10]. A weighted correlation kernel-based convolutional neural network is built for learning the hierarchical features [31]. In addition to CNNs, graph neural networks (GNNs) have gained prominence. GNNs have the capacity to capture information about neighboring structures within the brain. BrainGNN, for instance, introduced ROI-aware graph convolutional layers and ROI-selection pooling layers to predict neurological biomarkers at both the group and individual levels [13]. Another approach, proposed by [32], involved learning a graph similarity metric using a siamese graph convolutional neural network. A dynamic graph network is proposed by learning from sparse connections among brain regions calculated dynamically from graph features [12]. [33] proposed to perform a two-layer convolution on the fMRI and DTI data simultaneously. M-GCN regularized convolution on functional connectivity with structural graph Laplacian [34]. Cross-GNN is proposed to capture inter-modal dependencies [14]. Another type of GNN build transductive graphs and implement semi-supervised learning to predict via a node classification task [35]–[38].

More recently, the Transformer architecture has garnered considerable attention due to its exceptional performance in graph representation learning. However, most existing Transformer-based networks [39], [40] have achieved only limited success in brain network analysis. To address this limitation, BrainNetTransformer [16] was introduced with distinguishable cluster-aware embeddings to determine similar behaviors, which outperforms most of existing studies. However, these task-specific models are limited by the number of annotated samples, which limits their adaptation and generalization abilities to other tasks.

B. Self-supervised Learning

Self-supervised learning paradigms have delivered promising results in computer vision [22], [23] and natural language

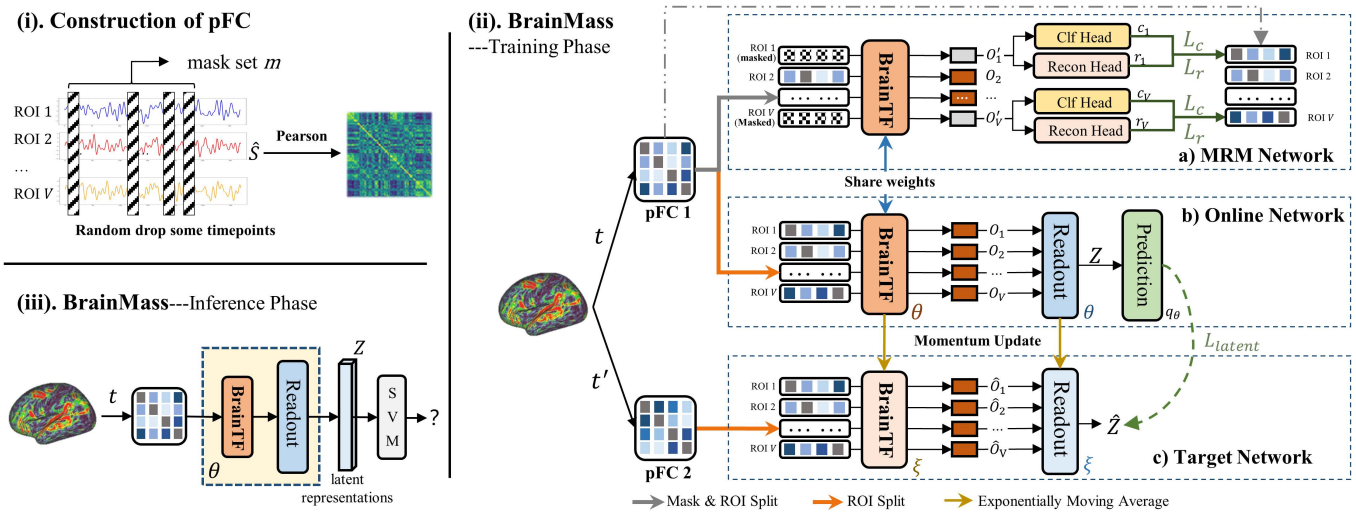

Fig. 1: Illustration of (i) the construction of pFC, (ii) the training phase of BrainMass method, including an MRM (an MRM network) and an LRA (an online network and a target network) module, and (iii) the inference phase of BrainMass.

processing [24], [25]. These paradigms have introduced pre-trained foundation models that leverage self-supervised learning on extensive unannotated data. This approach produces standardized and generalized representations, offering substantial benefits across domains with limited task-specific data availability.

In recent years, medical image analysis has achieved significant advancements through the use of foundation models and self-supervised learning (SSL) [26], [41]–[43]. For instance, BrainLM [44] is a Brain Language Model for brain activity dynamics analysis, which allows for the accurate prediction of clinical phenotypes as well as forecasting of future brain states. [45] pre-train the brain recordings based on 11,980 experimental runs of 1,726 individuals and adapt the pre-trained models to benchmark mental state decoding datasets. NeuroVNN is proposed as a foundation model for brain age prediction [46], facilitating extracting biologically plausible brain age estimates in different populations. Besides the brain signals, there are also several studies on other kinds of medical images. [47] introduces a novel approach utilizing multi-instance contrastive learning for X-ray classification, resulting in a 6.7% improvement in accuracy. [48] proposes a foundation model tailored for endoscopy video data, which outperforms state-of-the-art methods across a variety of tasks including classification, segmentation, and detection. Furthermore, [49] introduces a knowledge-enhanced auto-diagnosis system for analyzing paired chest X-rays and radiology reports, demonstrating exceptional zero-shot and few-shot performance that surpasses that of three expert radiologists on average. [50] presents RETFound, a foundation model designed for retinal images, which consistently exceeds the performance of several benchmark models in diagnosing and predicting sight-threatening eye diseases. Additionally, [51] introduces a visual-language foundation model for pathology image analysis, offering a generalizable solution that enhances model performance and reduces the annotation workload for experts, thereby facilitating broader clinical AI applications. However, most of these existing methods focus on images such as X-rays, retinal images, and endoscopy images, with a notable lack of research into applying self-supervised learning and foundation models to brain networks. BrainNPT [27], for example, constructs disturbance inputs by replacing regional features to enhance the models' understanding of the underlying input patterns. [19] leverages a masked seed-based strategy for pretraining. BrainGSLs [18], similarly, proposes an ensemble masked graph self-supervised framework based on masking and prediction. Nevertheless, these studies have only achieved modest improvements when compared to approaches without pre-training (i.e., approximately 71.5% accuracy on the ABIDE dataset). It's important to note that these pre-training strategies, borrowed from BERT-like models, still rely on a substantial amount of training data to establish data dependencies. Nonetheless, graph neural networks are limited by the depth of the models to be applied for the foundation model studies, due to the oversmoothing issue [52], [53]. In summary, there remains a notable gap in the development of self-supervised learning studies tailored to uncover the intrinsic characteristics of brain network.

III. Method

A. Preliminaries

The brain functional networks $\boldsymbol{X}$ are derived by mapping processed neuroimages onto a template with V Regions of Interest (ROIs). These networks are symmetric positive definite matrices, $\boldsymbol{X} \in \mathbb{R}^{V \times V}$. For diagnosis purposes, the goal is to develop a mapping function $f : \boldsymbol{X} \rightarrow y$, where y represents the predicted diagnosis phenotype for each subject.

In this study, we first generate two pFCs for each participant, and feed them into the BrainMass framework for pre-training a brain network Transformer (BrainTF) encoder. During the downstream classification phase, we froze the BrainTF and use it to extract latent representations, $\boldsymbol{Z}$, for each participant. The learned latent representations are further fed into a Support Vector Machine (SVM) classifier for downstream

prediction. This process is shown in Fig. 1. To note that, in the training phase, the BrainMass consists of three components: the MRM network, the online network, and the target network. Each network features a BrainTF encoder, sharing the same architectural design. The BrainTFs in the MRM and online networks share the same weights, while the BrainTF in the target network is updated by an exponential moving average based on the online network.

B. Pseudo functional connectivity augmentation

In this study, we propose to investigate brain network augmentation methods by random removal of certain timepoints within timeseries data. Consider a timeseries matrix $\boldsymbol{S} \in \mathbb{R}^{V \times T}$, where T represents the number of time steps. We generate a random dropping vector $\boldsymbol{m} \in \mathbb{R}^{M}$, with $M \leq T$, and use this vector to omit selected timepoints from the data, which is shown in Fig. 1 (i). Subsequently, we apply the Pearson correlation to the resulting modified timeseries matrix $\hat{\boldsymbol{S}} \in \mathbb{R}^{V \times (T-M)}$ to create a pseudo-functional connectivity (pFC) matrix $\hat{\boldsymbol{X}}$. It is important to note that fMRI data is often collected using a variety of protocols, resulting in differences in lengths of the timeseries data across samples. To address these variations, we investigate the effects of using different percentages of dropping lengths to account for the inherent variability in fMRI data acquisition and enhances the robustness and applicability of the augmentation.

C. Brain network Transformer encoder

Transformer-based models have led a tremendous success in various downstream tasks across fields including natural language processing, computer vision, and also graph learning. However, the brain networks potentially fall in neither of these classes. The brain networks are symmetric semi-positive defined matrices and densely distributed. Previous studies tackle the brain network as graph data, however, there are still no explicit inter-ROI relationships [13], [16], [37]. In this study, we instead tackle the connection profile as a sequence, where each ROI is represented as a sequential step with V features. The input brain network $\boldsymbol{X}$ is viewed as a sequence $\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_V\}$, where the i-th element is obtained by $\boldsymbol{x}_i = \boldsymbol{X}_{i,:} \in \mathbb{R}^V$. In this context, Multi-Head Self-Attention (MHSA) is implemented to relate inter-ROI dependencies and generate more expressive brain features:

$$\boldsymbol{H}^L = \text{MHSA}(\boldsymbol{X}) \in \mathbb{R}^{V \times V} \tag{1}$$

For each layer l, we first calculate the query $\boldsymbol{Q}^{l,c}$, key $\boldsymbol{K}^{l,c}$, and value $\boldsymbol{V}^{l,c}$ for the c-th head through linear projection as:

$$\boldsymbol{Q}^{l,c} = \boldsymbol{H}^{l-1}\boldsymbol{W}_q^{l,c}, \tag{2}$$

$$\boldsymbol{K}^{l,c} = \boldsymbol{H}^{l-1}\boldsymbol{W}_k^{l,c}, \tag{3}$$

$$\boldsymbol{V}^{l,c} = \boldsymbol{H}^{l-1}\boldsymbol{W}_v^{l,c} \tag{4}$$

where $\boldsymbol{H}^{l-1}$ is the output of the l-th layer, $\boldsymbol{H}^0 = \boldsymbol{X}$, and $W_q^{l,c}, W_k^{l,c}, W_v^{l,c}$ are learnable parameters. c is in the range of $\{1, 2, ..., C\}$, and C denotes the number of self-attention heads. The output for each head is computed as:

$$\boldsymbol{H}^{l,c} = \text{Softmax}(\frac{\boldsymbol{Q}^{l,c}(\boldsymbol{K}^{l,c})^T}{\sqrt{d}})\boldsymbol{V}^{l,c} \tag{5}$$

where d is the first dimension of $\boldsymbol{W}^{l,c}$. Finally, the output $\boldsymbol{H}^l$ is obtained by $\boldsymbol{H}^l = (||_{c=1}^{C}\boldsymbol{H}^{l,c})\boldsymbol{W}_O^l$, where $||$ is the concatenation operator, and $\boldsymbol{W}_O^l$ are learnable model parameters. We implement the Feed Foward Network and layer normalization for mapping the $\boldsymbol{H}^l$ into encoder outputs.

D. Masked ROI Modelling

For the mask modeling, as illustrated in the MRM network of the Figure 1 (ii), each input brain network is treated as a sequence, divided into V ROIs, and randomly assigned a set $\mathbb{P}$ of P masked ROI position indices. For each patch that needs to be masked, we replace its patch embedding with a learnable mask embedding. Positional embeddings are added to the patch embeddings, and the resulting data is fed into the BrainTF encoder. Classification (Clf) and Reconstruction (Recon) heads are utilized to predict the indices and features of the masked ROIs from the remaining ROI features.

For each masked patch $\boldsymbol{x}_i^{'}$, we obtain a corresponding output $\boldsymbol{o}_i^{'}$ from the BrainTF encoder. Subsequently, we pass $\boldsymbol{o}_i$ through both a classification head and a reconstruction head to obtain outputs $\boldsymbol{c}_i$ and $\boldsymbol{r}_i$, respectively. Both the classification and reconstruction heads consist of two-layer MLPs designed to map $\boldsymbol{o}_i^{'}$ to the same dimension as $\boldsymbol{x}_i^{'}$. The goal is to make $\boldsymbol{r}_i$ as close as possible to $\boldsymbol{x}_i^{'}$ while ensuring the model can correctly match pairs $(\boldsymbol{x}_i^{'}, \boldsymbol{c}_i)$. To achieve this, we employ the InfoNCE loss [54] L_c for the classification objective and the mean square error loss L_r for the reconstruction objective:

$$L_c = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{\exp(\boldsymbol{c}_i^T\boldsymbol{x}_i^{'})}{\sum_{j=1}^{N}\exp(\boldsymbol{c}_i^T\boldsymbol{x}_j^{'})}\right) \tag{6}$$

$$L_r = \frac{1}{N}\sum_{i=1}^{N}||\boldsymbol{r}_i - \boldsymbol{x}_i^{'}||^2 \tag{7}$$

E. Latent representation alignment

BrainMass leverages the latent representation alignment on two pFCs to achieve similar latent representations. Following previous works [23], [78], we employ a dual branch including the online network and the target network. The online network encodes brain networks using an L-layer Multi-head Self-Attention (MHSA) Transformer network, resulting in nonlinear mappings as $\boldsymbol{X} \rightarrow \boldsymbol{O} \in \mathbb{R}^{V \times V}$. A readout function subsequently transforms the encoded features $\boldsymbol{O}$ into embeddings $\boldsymbol{Z} \in \mathbb{R}^{D \times V}$. Similarly, the target network generates embeddings $\hat{\boldsymbol{Z}}$ through the same process. The target network provides regression targets for training the online network. To prevent model collapse, a predictor, q_θ, is used to maintain asymmetry between the online and target networks. A prediction Multi-Layer Perceptron (MLP) is employed to learn the mapping from the outputs $\boldsymbol{Z}$ of the online network to predict the outputs $\hat{\boldsymbol{Z}}$ of the target network. Parameters of the target network ξ are updated using an exponentially weighted moving average based on the online parameters θ:

$$\xi \leftarrow \tau\xi + (1-\tau)\theta \tag{8}$$

TABLE I: Demographical information on 30 datasets.

ID	Type for use	Dataset	Participants	Samples	Age (Mean/Std)	Gender (M/F)	Groups	Pretrain Samples	Downstream Samples
1	Type-I	UKBiobank [55]	21240	21240	54.90/7.49	10069/11171	Multiple	21240	-
2		CMI-HBN [56]	2228	4134	10.35/4.11	1428/800	Multiple	4134	-
3		GSP [57]	1570	2708	21.42/2.89	665/905	-	2708	-
4		CORR [58]	1515	4247	25.85/15.38	756/759	-	4247	-
5		NKI-RS [59]	1319	6863	39.00/22.07	515/804	-	6863	-
6		HCP [60]	1206	4176	(20-40)	550/656	-	4176	-
7		QTIM [61]	1202	1202	21.17/4.03	470/732	-	1202	-
8		FCON-1000	1070	1118	28.64/13.49	490/580	-	1118	-
9		SLIM [62]	1008	1008	20.01/1.20	451/557	-	1008	-
10		CCNP [63]	878	1581	11.15/3.04	469/409	-	1581	-
11		CAM-CAN [64]	652	652	54.85/18.54	322/330	-	652	-
12		CHINA-Project	608	608	62.48/9.98	308/300	AD, PD	608	-
13		SALD [62]	493	493	45.15/17.43	185/308	-	493	-
14		INDI-Retro	479	1494	33.19/16.82	239/240	-	1474	-
15		QTAB [65]	417	1142	11.56/1.65	215/202	-	1142	-
16		NAD [66]	300	1761	40.94/23.08	132/168	-	1761	-
17		ISYB [67]	215	215	22.60/2.66	59/156	-	215	-
18		CogTrain [68]	166	210	24.52/4.49	98/68	-	210	-
19		Caltech Conte Center [69]	117	305	28.53/6.41	68/49	-	305	-
20		Synaesthesia [70]	126	505	35.64/13.35	28/98	-	505	-
21	Type-II (Internal)	REST-MDD [71]	2379	2379	36.20/15.11	925/1454	MDD	1666	1276 MDD, 1104 NC
22		ADHD-200	1258	1258	11.72/3.29	765/493	ADHD	882	548 ADHD, 710 NC
23		OASIS [72]	1185	4516	70.06/9.43	446/739	DM	2222	309 DM, 300 NC
24		ADNI [73]	1171	3050	71.25/7.09	560/611	MCI, AD	2529	151 MCI, 149 AD, 142 NC
25		ABIDE-I	1114	1114	16.86/8.06	957/157	ASD	779	528 ASD, 556 NC
26		PPMI [74]	941	973	65.15/8.51	548/393	PD	864	70 PD, 70 proPD, 64 NC
27	Type-III (External)	ABIDE-II	1236	1236	14.66/9.12	895/341	ASD	0	559 ASD, 677 NC
28		LA5c [75]	192	192	34.04/9.32	118/74	Multiple	0	43 ADHD, 49 BP, 50 SCZ, 50 NC
29		Xuanwu [76]	213	213	61.72/9.61	107/106	PD, iRBD	0	90 PD, 53 iRBD, 70 NC
30		SchizoConnect [77]	188	188	37.90/12.76	143/45	SCZ	0	97 SCZ, 91 NC
		Total	46686	70781		23018/23748		64584	

NC: Normal Control, ASD: Autism Spectrum Disorder, ADHD: Attention Deficit Hyperactivity Disorder, DM: Dementia, PD: Parkinson's Disease, proPD: Prodromal Parkinson's Disease, BP: Bipolar Disorder, iRBD: Idiopathic Rapid Eye Movement Sleep Behavior Disorder, MDD: Major Depression Disorder, MCI: Mild Cognitive Impairment, AD: Alzheimer's Disease, SCZ: Schizophrenia

ADHD-200: `https://fcon_1000.projects.nitrc.org/indi/adhd200/`, INDI-Retro: `https://fcon_1000.projects.nitrc.org/indi/IndiRetro.html`, FCON-1000: `https://fcon_1000.projects.nitrc.org/fcpClassic/FcpTable.html`, ABIDE: `https://fcon_1000.projects.nitrc.org/indi/abide`.

where τ is a target decay rate $\tau \in [0, 1]$. The optimization is performed using the mean squared error between the normalized predictions and the target projections:

$$L_{latent} = 2 - 2\frac{< q_\theta(\boldsymbol{Z}), \hat{\boldsymbol{Z}} >}{||q_\theta(\boldsymbol{Z})||_2 \cdot ||\hat{\boldsymbol{Z}}||_2} \tag{9}$$

Readout function. After obtaining the non-linear features from the Transformer encoders, we are left with high-dimensional data, which pose challenges for downstream classification tasks, especially given the limited fMRI data samples available. In this study, we employ a readout function to transform the output features $\boldsymbol{O}$ into dimension-reduced embeddings. To achieve this, we aggregate the output features for each ROI into a set of D features. These features are then concatenated to form the final feature representation $\boldsymbol{O} \in \mathbb{R}^{D \times V}$. In this study, we set $D = 8$ and finally obtain $8 \times V$ representations for each brain network.

F. Optimization

To finalize the objective function for self-supervised pre-training, we employ a weighted summation that includes balancing parameters λ_c and λ_r as:

$$L = L_{latent} + \lambda_c L_c + \lambda_r L_r \tag{10}$$

This regularization enables the brainTF encoder to learn intra-network dependencies across various brain connectome patterns, enhancing its effectiveness for downstream tasks.

IV. Experiments

Datasets. We built a large cohort for both pretraining and evaluation, with detailed demographic information presented in Table I. This cohort includes participants from multiple centers, races, and countries throughout the world. To the best of our knowledge, this is the largest dataset for brain network analysis, featuring a diverse group of participants comprising a range of diagnosis types, including Normal Control (NC) subjects and individuals with various neurodevelopmental and neurodegenerative disorders. These disorders encompass Major Depression Disorder (MDD), Mild Cognitive Impairment (MCI), Alzheimer's Disease (AD), Autism Spectrum Disorder (ASD), Attention Deficit Hyperactivity Disorder (ADHD), Dementia (DM), Parkinson's Disease (PD), Prodromal PD (proPD), Bipolar Disorder (BP), and Idiopathic Rapid Eye Movement Sleep Behavior Disorder (iRBD).

The datasets are categorized into three types based on their utilization: (1) Type-I Datasets (Pre-training): 20 datasets are reserved for pre-training purposes. (2) Type-II Datasets (Pre-training and internal evaluation): Six datasets are used both for pre-training and evaluation. From these, 70% of samples are randomly selected and set fixed use for pre-training and downstream training. (3) Type-III Datasets (External evaluation): Four datasets are designated as external datasets. These are crucial for assessing the model's generalization and adaptability capabilities. A total of 64,584 samples are incorporated into the pre-training phase. Ten datasets are specifically employed to evaluate the diagnosis of brain diseases. Notably, for fair

comparison, single scan of fMRI for each subject was collected in the evaluation, and the duplicated scans are excluded for both pretraining and downstream evaluations.

Evaluation. In this study, to ensure a fair comparison, duplicated scans from Type-II and Type-III datasets have been excluded for the downstream tasks.

REST-MDD, ADHD-200, ABIDE-I, ABIDE-II, LA5c, SchizoConnect, and Xuanwu datasets. These datasets comprise participants, each with a single scan. The samples are randomly assigned with 70% used for training and the remaining 30% for validation and evaluation.

Xuanwu. A total of 213 subjects are included in this dataset, comprising 70 Normal Controls (NCs), 53 subjects with idiopathic Rapid Eye Movement Sleep Behavior Disorder (iRBD), and 90 subjects with Parkinson's Disease (PD). These individuals were recruited from the Movement Disorders Clinic of Xuanwu Hospital at Capital Medical University. All participants provided written informed consent for the experiment. The iRBD patients were screened using the International Classification of Sleep Disorders-Third Edition (ICSD-3) diagnostic criteria and confirmed through polysomnography. The NCs were all older than 40 years, had no family history of movement disorders, and no significant cerebral lesions were observed in their MR images. The PD subjects were diagnosed according to the Movement Disorder Society's Clinical Diagnostic Criteria for Parkinson's Disease.

OASIS. The OASIS dataset is a longitudinal dataset, consisting of 1185 subjects and 4516 samples. Among these, 309 participants are diagnosed with dementia (DM), and 603 are normal controls. We selected 309 DM and 300 NC participants by matching for age and sex. For downstream evaluations, 609 samples are used. The remaining participants with 2222 samples are employed for pretraining purposes.

ADNI. The ADNI dataset is compiled from multiple subsets, including ADNI-1, ADNI-Go/2, and ADNI-3. It is important to note that the ADNI dataset is longitudinal, with each participant undergoing multiple scans. Among the ADNI participants, 164 who have undergone fMRI scans were either diagnosed with Alzheimer's Disease (AD) or later converted to AD. To complement this, 162 NCs and 161 participants with Mild Cognitive Impairment (MCI) were included, matched for age and sex. After image preprocessing and quality control, 161 MCI cases, 149 AD cases, and 162 normal controls were selected. The remaining 2529 samples from 699 subjects were used for pretraining.

PPMI. The PPMI dataset comprises 316 prodromal participants, 64 normal controls, and 322 individuals diagnosed with Parkinson's Disease, all of whom have undergone fMRI scans. To align with the 64 normal controls, we selected 70 individuals with PD and 70 prodromal PDs (proPDs) for comparison. The diagnostic criteria for PD adhered to the inclusion criteria for patients in the PPMI study.

Data Preprocessing. All the fMRI images were preprocessed by reference to the Configurable Pipeline for the Analysis of Connectomes (C-PAC) pipeline , including skull striping, slice timing correction, motion correction, global mean intensity normalization, nuisance signal regression with 24 motion parameters, and band-pass filtering (0.01-0.08Hz). The functional images were finally registered into standard anatomical space (MNI152). The mean time series for a set of regions were computed and normalized into zero mean and unit variance. Pearson Coefficient Correlation was applied to measure functional connectivity. In this study, the preprocessed fMRI images were mapped by the brain template for parcellations by the Schaefer atlas [79] into 100 ROIs.

Implementation details. For pretraining, we utilized the Adam optimizer with an initial learning rate of 3×10^{-5} and a weight decay of 5×10^{-5}. The learning rate underwent a linear increase to 3×10^{-4} within 10 warmup epochs. The batch size was set fixed as 256. The BrainMass model underwent training for 2000 epochs, and we saved the models with the lowest training loss for subsequent classification tasks. The decay rate τ for target network update is set as 0.996. Our experiments were conducted on a platform equipped with 64 NVIDIA Tesla V100 GPUs, with 8 GPUs allocated for each training. It takes around 150 hours for each pretraining. The implemented Transformer encoder is configured with 32 layers, and 20 heads for MHSA. The hidden feature dim is 4096 for FFN. The total parameter size is 67.0M. For the optimization, we set λ_c and λ_r to 0.1 and 5 in Eq. 10. In the downstream tasks, the latent representations were input into an SVM classifier for prediction. To facilitate a more robust comparison on smaller-sized datasets, we repeated the downstream tasks 10 times by randomly sampling the validation and test sets.

Metrics. We assess the performance of diagnosis classification using accuracy (ACC), sensitivity (SEN), and specificity (SPE) as our key metrics. We employ a rigorous stratified sampling strategy that considers collection sites during the training (70%)-validation (15%)-testing (15%) split, ensuring fair comparisons [16].

V. Results

A. Brain disorder diagnosis performance

For comparison, two categories of baseline models are included: those with SSL and those without SSL. The baseline models without SSL include BrainNetCNN [9], DHGNN [80], BrainGNN [13], Semi-GCN [35], vanilla-Transformer (vanillaTF), and BrainNetTransformer (BrainNetTF) [16]. For SSL comparisons, powerful SSL frameworks like BYOL [78] and MOCO [22] are included. Furthermore, we considered two existing works: BrainNPT [27] and BrainGSLs [18].

Table II presents the results from 8 tasks across 6 internal datasets, with the highest performance marked in bold and the second-best underlined. Notably, for the REST-MDD dataset, as raw images were not available, we additionally trained a model by mapping fMRIs with the AAL atlas into 116 ROIs. From these results, we observe the following key points:1) Among these models, CNN, GNN, and Transformer architectures demonstrate similar performance across all tasks. Despite their increased computational complexity, Transformer models offer limited performance enhancement when handling fMRI data with limited sample sizes. However, BrainNetTF significantly boosts diagnostic performance, corroborating findings from previous studies [16]. 2) Contrary to expectations, most SSL approaches did not markedly enhance performance in

TABLE II: Classification results of different approaches on 8 tasks of 6 internal datasets in terms of accuracy (Acc), sensitivity (Sen), and specificity (Spe). SSL indicates the model is pretrained by self-supervised learning.

Dataset	SSL	ABIDE-I			ADHD-200			REST-MDD*			OASIS		
Task		NC vs. ASD			NC vs. ADHD			NC vs. MDD			NC vs. DM		
Metric		ACC ↑	SEN ↑	SPE ↑	ACC ↑	SEN ↑	SPE ↑	ACC ↑	SEN ↑	SPE ↑	ACC ↑	SEN ↑	SPE ↑
BrainNetCNN		68.14-2.04	67.56-7.02	69.74-4.41	61.62-1.81	63.18-8.39	61.82-1.46	62.55-1.81	64.41-8.39	59.93-1.46	68.24-3.66	67.42-4.21	69.53-3.22
DHGNN		64.31-1.52	63.81-5.03	64.97-2.62	59.84-2.04	53.52-2.65	61.72-2.51	59.24-2.04	61.40-2.65	56.51-2.51	66.71-5.61	66.39-4.90	67.29-6.88
BrainGNN		69.60-2.24	61.47-3.59	71.46-2.57	61.02-2.59	54.60-4.05	64.08-2.85	61.40-2.59	61.37-4.05	55.86-2.85	68.24-2.27	61.74-9.29	74.89-7.31
PopGCN		69.76-1.40	67.61-3.12	71.72-1.74	62.20-1.36	56.69-2.17	**66.40-2.98**	61.20-1.36	64.06-2.17	57.23-2.98	65.93-3.64	66.00-4.35	65.82-3.24
vanillaTF		68.98-1.13	65.01-5.19	**72.48-4.03**	61.62-1.14	63.18-5.20	61.82-1.74	62.49-1.14	**64.61-5.20**	60.65-1.74	68.57-2.26	68.38-3.80	69.74-3.31
BrainNetTF		71.02-1.16	73.27-5.62	71.18-4.38	62.75-1.29	63.61-5.25	62.85-2.25	63.50-1.14	64.05-5.20	61.56-1.74	72.53-2.41	74.55-5.58	71.41-2.26
MoCo	✓	68.68-2.50	66.01-2.97	70.85-3.45	58.95-2.77	51.02-6.25	62.66-2.89	62.34-1.55	61.64-1.12	63.83-2.76	71.32-3.88	69.48-3.71	74.18-5.57
BYOL	✓	68.98-1.49	67.88-3.75	70.00-2.82	59.46-3.55	51.98-8.60	63.12-2.28	62.80-1.16	62.43-0.82	63.61-2.15	68.79-3.97	70.17-4.66	67.70-3.79
BrainNPT	✓	63.83-2.84	61.51-4.32	65.51-2.99	58.27-2.52	55.77-9.39	58.68-2.24	57.84-1.31	58.84-0.97	56.01-2.03	64.51-2.74	65.02-3.58	65.01-4.15
BrainGSLs	✓	70.98-3.68	70.62-4.10	71.71-4.94	62.32-2.96	61.48-5.28	66.25-5.42	59.87-2.67	59.11-3.10	62.22-1.23	59.87-2.67	59.11-3.10	62.22-1.23
BrainMass	✓	**72.75-2.85**	**74.26-3.88**	71.89-5.13	**64.65-1.66**	**64.36-6.16**	64.85-1.38	**66.42-1.16**	**64.61-1.24**	**67.61-1.41**	**76.48-2.26**	**75.14-2.08**	**78.25-3.53**
Dataset	SSL	ADNI			ADNI			PPMI			PPMI		
Task		NC vs. MCI			NC vs. AD			NC vs. proPD			NC vs. PD		
Metric		ACC ↑	SEN ↑	SPE ↑	ACC ↑	SEN ↑	SPE ↑	ACC ↑	SEN ↑	SPE ↑	ACC ↑	SEN ↑	SPE ↑
BrainNetCNN		56.51-3.90	59.12-6.76	55.61-5.01	67.21-4.34	69.76-7.54	67.69-5.63	69.44-6.69	62.22-6.78	81.69-8.80	73.00-6.40	70.24-6.62	77.56-7.02
DHGNN		52.09-4.90	53.02-4.07	48.43-1.72	61.40-4.19	63.45-7.01	62.51-5.97	71.67-3.89	63.95-4.66	86.18-8.08	70.50-5.68	68.38-4.99	77.44-10.24
BrainGNN		61.16-5.61	55.46-9.04	**67.14-10.31**	71.63-3.42	74.55-8.18	68.57-10.09	67.22-5.58	63.89-8.29	84.89-3.46	76.50-4.50	79.00-5.39	64.00-11.14
PopGCN		61.43-7.00	61.61-7.17	60.84-7.57	66.16-5.76	65.56-6.60	66.15-6.04	67.78-5.24	62.98-5.90	73.44-7.81	67.50-6.80	66.97-7.77	67.48-6.91
vanillaTF		57.91-5.45	60.03-6.21	57.54-7.09	72.56-4.51	69.84-6.14	78.48-5.90	72.22-3.51	64.41-3.73	84.60-6.67	74.00-7.68	70.87-7.59	81.11-11.42
BrainNetTF		62.10-3.46	64.53-4.70	61.44-4.67	75.81-3.78	81.05-9.44	73.77-4.68	72.22-4.97	65.76-7.64	83.00-6.81	77.00-9.27	73.95-9.64	82.53-9.43
MoCo	✓	60.70-5.14	59.46-4.16	63.19-7.16	74.42-5.79	75.29-6.03	74.26-7.19	77.22-7.64	73.20-9.07	83.53-9.11	83.50-4.50	70.64-5.14	**88.68-8.36**
BYOL	✓	56.74-3.16	57.32-2.87	56.21-3.83	74.93-5.91	74.63-5.30	75.80-7.46	77.78-8.96	72.52-9.38	86.88-9.65	64.00-7.00	61.22-6.01	69.67-9.76
BrainNPT	✓	55.58-4.09	55.25-3.44	56.21-5.50	66.98-3.72	68.84-2.46	59.72-8.60	72.78-8.33	69.54-9.58	76.81-11.18	61.50-4.50	61.00-3.00	65.50-2.55
BrainGSLs	✓	57.33-7.10	60.23-7.86	60.27-7.08	76.64-3.28	81.12-8.78	**82.88-5.36**	76.35-7.86	73.52-9.03	87.46-4.61	77.53-8.56	78.38-9.18	79.97-8.68
BrainMass	✓	**68.37-3.48**	**70.51-5.07**	66.61-2.71	**83.72-2.55**	**85.94-5.86**	82.29-2.79	**80.00-4.44**	**74.70-9.29**	**88.18-8.66**	**84.50-6.87**	**82.69-7.63**	87.31-7.57

TABLE III: Ablation studies on the elements of BrainMass with the accuracy (%) performance on eight internal tasks.

L_{Latent}	MRM		ABIDE-I	ADHD-200	REST-MDD	OASIS	ADNI		PPMI	
	L_c	L_r	ASD	ADHD	MDD	DM	MCI	AD	proPD	PD
✓			63.83-2.84	58.27-2.52	62.80-1.16	64.51-2.74	53.26-4.47	70.93-5.91	77.78-8.96	64.00-7.00
		✓	65.39-3.01	60.76-1.90	64.65-2.67	69.01-4.28	55.12-4.54	66.05-5.32	77.78-3.51	74.00-7.68
✓	✓		67.13-2.48	60.16-1.43	62.46-2.44	67.98-2.66	61.40-3.15	72.09-3.75	74.44-6.67	**84.50-6.50**
✓		✓	71.62-2.84	61.62-2.84	66.00-1.31	71.98-2.96	61.86-4.67	79.07-2.55	**80.00-3.68**	**84.50-6.50**
✓	✓	✓	**72.75-2.85**	**64.65-1.66**	**66.42-1.16**	**76.48-2.26**	**68.37-3.48**	**83.72-2.55**	**80.00-4.44**	**84.50-6.87**

comparison to BrainNetTF. In fact, some SSL methods even underperformed relative to baseline models without SSL. 3) Our BrainMass model consistently outperforms these methods across all 8 tasks, with accuracy improvements of 1.73%, 2.33%, 2.92%, 3.95%, 6.27%, 7.08%, 2.22%, and 1.00% for distinguishing ASD, ADHD, MDD, DM, MCI, AD, proPD, and PD, respectively. This highlights the significant benefits of large-scale SSL representations and underscores the effectiveness of our BrainMass framework.

Additionally, while previous studies [35], [81] demonstrate higher accuracy in distinguishing AD/MCI from NC, our results indicate lower overall accuracy. On one hand, our study utilizes a single scan per participant, unlike others that incorporate multiple scans to expand the dataset. Repeated scans might exhibit similar functional features, boosting accuracy. Compared with them, in our setting, multimodal approaches could only enhance the accuracy to 88.6% in our previous studies [14]. On the other hand, we adopt a rigorous model selection strategy by choosing the optimal model based on the validation set, aligning with [16], [35], which might affect generalized performance due to the validation-test gap.

B. Sensitive analysis and ablation studies

Ablation studies. We undertook evaluations focusing on the distinct components of Eq. 10. These components include latent representation learning (L_{latent}), as well as MRM with classification heads (L_c) and reconstruction heads (L_r) for distinguishing masked ROI indices and features. The results corresponding to these evaluations are compiled in Table III. We can find that incorporating the L_r term significantly enhances performance across all tasks. While the ROI meta-label prediction term (L_c) has a comparatively modest impact on its own, its integration with $L_{latent} + L_r$ yields further improvements in model performance. This enhancement can be attributed to the model's increased proficiency in capturing dependencies among brain regions and in learning intra-network representations. The masked ROI distinguishing module plays a pivotal role in this process, facilitating a more nuanced and effective learning paradigm. When we finally combined all these elements, the performances are further improved in most cases.

Drop ratio analysis. In this study, we introduce the approach of constructing pFCs to generate millions of brain networks, with a specific focus on mitigating temporal dynamics. A critical aspect of this process is the dropping rate, which we identify as a key hyperparameter. To explore its impact, we analyzed the accuracy performance of our model across various dropping rates, ranging from 10% to 40%, as depicted in Figure 2 B). Our observations reveal an initial improvement in performance as the drop ratio increases, followed by a decline once a certain threshold is exceeded. We found that, across all tasks, the optimal dropping rate falls within the range of 10% to 20%. Notably, when the dropping ratio surpasses 40%, there is a consistent deterioration in

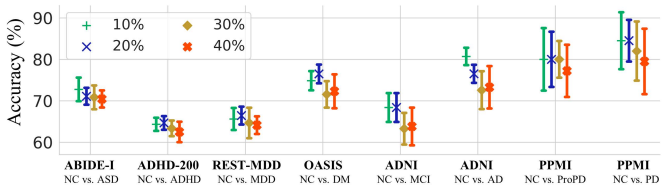

Fig. 2: The effect on the dropping rate on eight internal tasks.

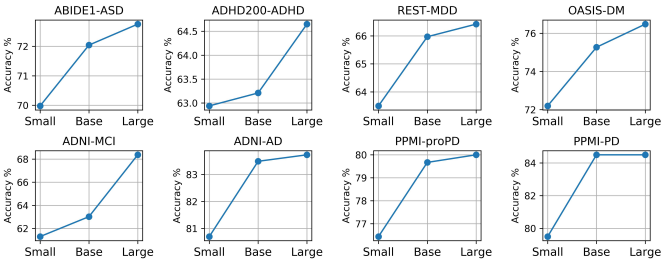

Fig. 3: The effect on the model size.

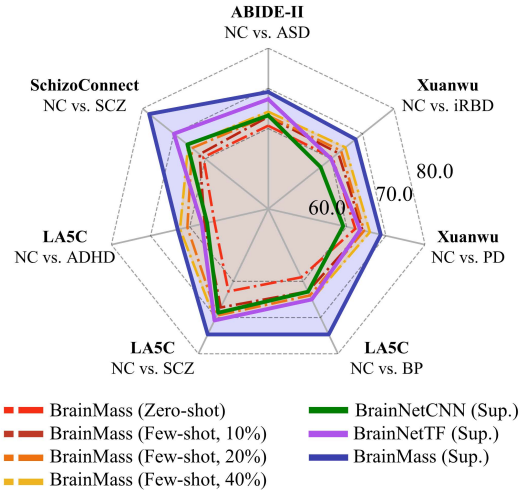

Fig. 4: The accuracy performances on seven external tasks.

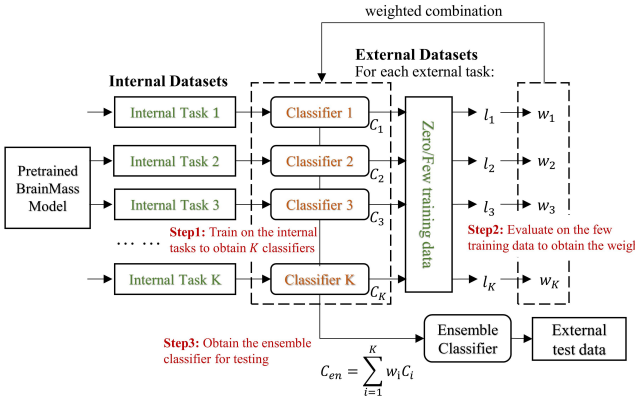

Fig. 5: The workflow of the zero/few-shot learning for BrainMass.

performance across all 8 tasks. This pattern highlights the delicate balance required in selecting the appropriate dropping rate for optimal model performance. It suggests that while some degree of dropping can enhance model efficacy by reducing temporal noise, excessive dropping may lead to the loss of critical temporal information, adversely affecting the model's ability to accurately analyze brain networks.

Model Size analysis. Figure 3 presents the accuracy of models of varying sizes, categorized as small, base, and large. These models are equipped with 8, 16, and 32 Transformer layers, 5, 10, and 20 attention heads, and 1024, 2048, and 4096 FFN features, respectively. Their total parameters amount to 14.4 M, 25.4 M, and 67.0 M, respectively. The results show that the large model configuration achieves superior performance in all tested scenarios. Furthermore, there is a noticeable trend of increasing accuracy as the model becomes larger. The base model demonstrates significant performance improvements in six tasks compared with the small model, except on the ADHD and MCI classification tasks. The large model not only enhances performance slightly across these six tasks but also shows significant improvements in ADHD and MCI classification tasks. Consequently, large models trained with more compute and parameters, exhibit potential emergent abilities with substantial performance increases.

C. Generalizability and few/zero-shot evaluation

We extended our evaluation to external datasets, benchmarking our BrainMass against the baseline BrainNetCNN and the state-of-the-art BrainNetTF. As shown in Figure 4, we analyzed the accuracy scores across seven tasks, involving distinguishing NC from ASD, SCZ, ADHD, BP, PD, and iRBD on ABIDE-II, SchizoConnect, LA5C, and Xuanwu datasets. Our BrainMass consistently outperformed both BrainNetCNN and BrainNetTF in these tasks, achieving improvements of 1.81%, 7.78%, 4.22%, 3.57%, 3.87%, 6.15%, and 7.86% over BrainNetTF, respectively.

Furthermore, our diagnostic task involved distinguishing disorders from NC. To this end, we further studied whether this could be generalized to other diseases. We developed several classifiers using internal datasets and applied ensemble learning for few-shot and zero-shot learning on external datasets. We illustrate the workflow of zero/few-shot learning in Figure 5. Specifically, we first obtain K classifiers for the internal tasks by framing the classification task as distinguishing abnormal from normal cases. Subsequently, we determine the ensemble weights w_i for each classifier C_i. For zero-shot inference, the ensemble weights are calculated by averaging the prediction probabilities of the classifiers, where $K = \frac{1}{K}$. For few-shot learning, a weighted summation is applied, based on the prediction accuracy error l_i of the available samples. More precisely, the weights are calculated as $w_i = \frac{\log(l_i)}{\sum_{j=1}^{K} \log(l_j)}$. Ultimately, we achieve an ensemble classifier for inferring on the test data. In this study, to maintain atlas consistency, we use the Schaefer atlas, resulting in $K = 7$ internal classifiers. The corresponding performances are shown in Fig. 4 with light red indicating zero-shot, and dark red, orange, and yellow for few-shot with 10%, 20%, 40% samples. From the resutls, we can see that zero-shot inference sometimes equals or exceeds the supervised baseline BrainNetCNN. Moreover, with 20% annotated samples, BrainMass even surpasses BrainNetTF in

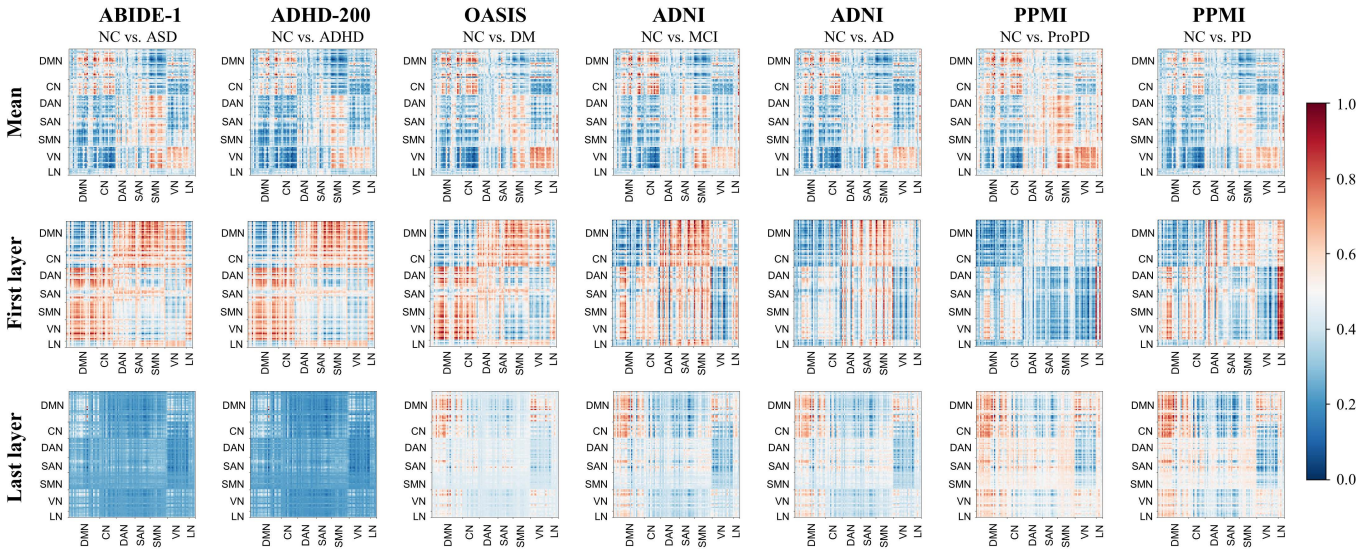

Fig. 6: Heatmaps of the Transformer encoder attention maps on 7 tasks, including the averaged attention maps (the first row), those of the first layer (the second row), and the last layer (the third row). The values in heatmaps are normalized into 0 to 1.

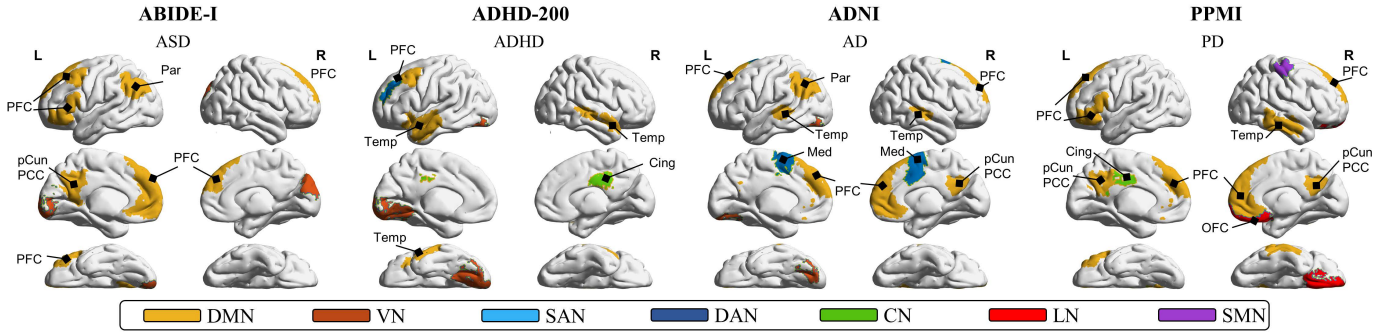

Fig. 7: Visualization on the ten key regions. The key regions are colored with the corresponding sub-network. Temp: the temporal. Par: the parietal. Cing: the cingulate. Med: the medial. PFC: the prefrontal cortex. pCun: the precuneus. PCC: the posterior cingulate cortex. OFC: the orbital frontal cortex.

differentiating ADHD and iRBD. This highlights BrainMass's remarkable capability in generalizing to various diseases, showcasing its potential for clinical applications with limited annotated samples.

D. Biological explanation

We present the attention maps across seven internal tasks in Figure 6, with the averaged heatmaps, the heatmaps of the first layer, and those of the last layer. Using the Schaefer atlas, the brain network is divided into seven sub-networks: the default mode network (DMN), the visual network (VN), the salience ventral attention network (SAN), the dorsal attention network (DAN), the control network (CN), the somatomotor network (SMN), and the limbic network (LN). From the first row, we can see that the BrainTF encoder generally obtains similar averaged heatmaps for different brain disorders. This insight is crucial for zero-shot/few-shot learning, as it provides explainable evidence on our model's ability to differentiate disorders from the normal.

In addition, across all diseases, we find that the shallower layers focus more on the interactions of DMN and CN with other networks, while this trend shifts in deeper layers. Inter-network communication is significant since the brain is not made up of isolated networks and many tasks require information passing and neuron firing through multiple networks [82]. This communication involves a complex balance among networks with profound implications for understanding human behavior in health and disease [83]. Our model exhibits this characteristic in low-level information processing within shallow layers, especially in interactions between the DMN and other networks, which are critical for cognitive control tasks like attention, memory, and execution. Conversely, at higher levels of information processing within deeper layers, the models discern more disease-specific patterns. We suspect that this process closely mirrors the mechanisms of the human brain. For a cognition task, the initial step involves interactions between different subnetworks, which are crucial for determining how information is processed and how the task is initiated. The focus then shifts to specific subnetworks that handle the intrinsic workload of the task. This aligns with research suggesting that cognitive processing involves dynamic and complex interplay among various brain regions,

TABLE IV: p-values after correction of the 10 key brain regions. * indicates the key regions with significant difference.

	ABIDE-I NC vs. ASD		ADHD-200 NC vs. ADHD		ADNI NC vs. AD		PPMI NC vs. PD	
ID	ROI	P-value (FDR)	ROI	P-value (FDR)	ROI	P-value (FDR)	ROI	P-value (FDR)
1	LH.DMN.pCunPCC_1	0.035*	LH.VN_5	0.010*	RH.SAN.Med_2	0.005*	RH.DMN.pCunPCC_2	0.026*
2	LH.DMN.PFC_6	0.040*	LH.SAN.PFCl_1	0.030*	RH.DMN.pCunPCC_2	0.015*	LH.DMN.pCunPCC_2	0.029*
3	LH.DMN.PFC_2	0.045*	RH.DMN.Temp_2	0.030*	LH.SAN.Med_3	0.024*	RH.DMN.PFCdPFCm_1	0.037*
4	LH.DMN.pCunPCC_2	0.080	LH.DMN.Temp_1	0.060	LH.DMN.Par_2	0.048*	RH.SMN.6	0.042*
5	LH.DMN.Par_2	0.100	RH.DN.Cing_1	0.065	LH.DMN.Temp_2	0.048*	RH.DMN.PFCdPFCm_2	0.062
6	LH.DMN.PFC_5	0.168	LH.DMN.PFC_6	0.094	LH.DMN.PFC_5	0.052	RH.DMN.Temp_1	0.068
7	LH.DMN.PFC_3	0.236	LH.VN_3	0.094	RH.DMN.Temp_3	0.076	RH.LN.OFC_1	0.094
8	RH.VN_8	0.236	RH.DMN.Temp_3	0.095	LH.VN_2	0.060	LH.DMN.PFC_5	0.084
9	LH.VN_5	0.248	LH.DMN.PFC_1	0.097	RH.DMN.PFCdPFCm_1	0.070	LH.DMN.PFC_2	0.096
10	RH.DMN.PFCdPFCm_2	0.275	LH.VN_2	0.097	RH.DMN.PFCdPFCm_2	0.070	LH.CN.Cing_1	0.099

LH: the left hemisphere. RH: the right hemisphere. Temp: the temporal. Par: the parietal. Cing: the cingulate. Med: the medial. PFC: the prefrontal cortex. pCun: the precuneus. PCC: the posterior cingulate cortex. OFC: the orbital frontal cortex.

each contributing uniquely to the cognitive function [84], [85].

In the last layer (the third row), we can see that the disease patterns bifurcate into two categories visually: neurodevelopmental (ASD and ADHD) and neurodegenerative diseases, each with similar patterns within their group. This observation demonstrates that our BrainMass has the abilities to interpret the differences between various diseases. To this end, we present the 10 key brain regions in Fig. 7, using multivariate analysis [76], [86]. These regions are selected based on their corresponding p-values after correction. The p-values are shown in Table IV, where an asterisk (*) indicates regions with significant differences. Consistently found across all diseases is the DMN, emerging as a critical network. It is associated with task-irrelevant mental processes, emotional and self-referential cognitive control, and memory encoding [87]. DMN alterations might contribute to attention lapses and memory deficits observed in AD, PD, ASD, and ADHD. Additionally, other crucial biomarker regions, such as the SMN in PD progression and the LN in AD progression, are also identified, consistent with previous studies [88], [89]. Overall, these findings suggest that our BrainMass allows for the meaningful interpretation of key biomarkers.

VI. Discussion

In this study, we introduce BrainMass, a self-supervised learning approach specifically tailored to the brain network analysis. Our study is bifurcated into two core parts: pseudo-functional connectivity (pFC) data augmentation and a BrainMass SSL framework. This framework is comprised of masked modeling and latent representation learning, both integral to the improvements of downstream classification tasks. This paradigm has yielded significant advancements. On one hand, it has notably improved disease diagnosis performance, with BrainMass demonstrating superior improvements over state-of-the-art (SOTA) models in both internal and external validations, underscoring its potent diagnostic capabilities and generalizable representations. On the other hand, neuroscience studies are currently grappling with challenges such as acquiring large-scale, hard-to-obtain datasets and dealing with high variability in scanning protocols. In this context, our pretrained models offer the versatility to generalize across diverse scenarios and help mitigate the risk of overfitting. Our pretrained model facilitates the extraction of relevant features and the application of an SVM for classification without the need for retraining or fine-tuning the encoders.

Besides, we found that the performances can be even further improved. For instance, pretraining on the ABIDE-I training set and evaluating on the remainder of the ABIDE-I dataset showed that accuracy could be further improved to 73.1%, surpassing all previous studies. Compared with the results in Table II, we observed that models pretrained on large-scale datasets exhibited slightly lower accuracy than those pretrained on a single dataset. This phenomenon is likely due to enhanced generalizability at the expense of decreased specificity for downstream classification. In addition, significant variability in data domains across different data centers might introduce domain noise, affecting downstream tasks. This highlights a trade-off between generalization and specificity. Our BrainMass model, pretrained on a wide range of datasets, demonstrates the ability to generalize across various tasks and datasets, while maintaining relatively promising performance.

Moreover, in this study, due to the imbalance between the downstream datasets and the pretrained model size, we leverage the SVM classifier for downstream evaluations, which shows powerful generalizability without end-to-end fine-tuning. Additionally, powerful fine-tuning tools such as LoRA [90] and DoLA [91] would be implemented to further improve performance and adaptability. One of our future works involves fine-tuning methods on small datasets with large models.

Finally, it's worth noting that brain network-based approaches, including our proposed BrainMass, are limited in accuracy and only show incremental improvements in accuracy for diagnosing some brain disorders. However, this limitation is primarily because functional changes provide limited insights into disease progression for some diseases. Structural changes, particularly in neurodegenerative disorders like AD and PD are also crucial diagnostic factors. In our future work, we aim to develop multi-modal neuroimaging pretrained encoders and establish inter-modal dependencies through mapping alignments. This approach will further enhance our capacity for brain disease diagnosis and the understanding of the brain disorder progression.

VII. Conclusion

In this study, we propose BrainMass, the first foundation model specifically designed for brain network analysis and disease diagnosis through functional measurements. BrainMass leverages the MRM and LRA modules to pre-train the Transformer encoder, focusing on intra-network dependencies and bootstrapped regularized latent representations. Our BrainMass model fosters generalizable and homogeneous representations, facilitating a wide range of brain disorder diagnoses using a single model set. Moreover, visualizations of the attention maps and multivariate analysis of the latent representations demonstrate the model's potential emergent ability to discriminate between abnormal and normal states. This highlights its potential for clinical application with robust zero-shot and few-shot learning capabilities. Our study provides new insights into the application of large-scale self-supervised learning in the realm of brain functional network analysis and addresses the lack of large models in brain network analysis.

VIII. Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 56113.

GSP data were provided by the Brain Genomics Superstruct Project of Harvard University and the Massachusetts General Hospital, with support from the Center for Brain Science Neuroinformatics Research Group, the Athinoula A. Martinos Center for Biomedical Imaging, and the Center for Human Genetic Research.

Data collection and sharing for the ADNI dataset was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

PPMI Data used in this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (`www.ppmi-info.org/access-dataspecimens/download-data`), RRID: SCR 006431.

HCP Data were provided by the Human Connectome Project, WU-Minn Consortium funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Data collection and sharing for SchizoConnect project was funded by NIMH cooperative agreement 1U01MH097435. SchizConnect Data used in this article were obtained from the SchizConnect database (`http://schizconnect.org/`). As such, the investigators within SchizConnect contributed to the design and implementation of SchizConnect and/or provided data but did not participate in the analysis or writing of this report.

Data used in the CAM-CAN project this work were obtained from the CamCAN repository. Data collection and sharing for this project was provided by the Cambridge Centre for Ageing and Neuroscience (CamCAN). CamCAN funding was provided by the UK Biotechnology and Biological Sciences Research Council (grant number BB/H008217/1), together with support from the UK Medical Research Council and University of Cambridge, UK.

The CHina Initiative on Neurodegeneration and Aging (CHINA) Project and Xuanwu datasets are curated by Xuanwu Hospital, Capital Medical University, China and were all approved by the Institutional Review Board of Xuanwu Hospital. They are funded by the National Key Research and Development Program of China (2021YFC2501202).

We extend our heartfelt gratitude to Dr. Tao Wu from Capital Medical University, Beijing, China, as well as Dr. Kunru Song and Dr. Jintao Zhang from the State Key Laboratory of Cognitive Neuroscience and Learning and the IDG/McGovern Institute for Brain Research at Beijing Normal University, Beijing, China. Their unwavering dedication and contributions were instrumental toward accomplishing our research objectives. Their insightful input and guidance during the data analysis phase greatly enriched our study.

References

[1] N. K. Logothetis, "What we can do and what we cannot do with fmri," *Nature*, vol. 453, no. 7197, pp. 869–878, 2008.

[2] D. J. Heeger and D. Ress, "What does fmri tell us about neuronal activity?" *Nature reviews neuroscience*, vol. 3, no. 2, pp. 142–151, 2002.

[3] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fmri signal," *nature*, vol. 412, no. 6843, pp. 150–157, 2001.

[4] A. Fornito, A. Zalesky, and M. Breakspear, "The connectomics of brain disorders," *Nature Reviews Neuroscience*, vol. 16, no. 3, pp. 159–172, 2015.

[5] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls," *Frontiers in systems neuroscience*, vol. 9, p. 175, 2016.

[6] N. J. Shah, A.-M. Oros-Peusquens, J. Arrubla, K. Zhang, T. Warbrick, J. Mauler, K. Vahedipour, S. Romanzetti, J. Felder, A. Celik *et al.*, "Advances in multimodal neuroimaging: hybrid mr–pet and mr–pet–eeg at 3 t and 9.4 t," *Journal of Magnetic Resonance*, vol. 229, pp. 101–115, 2013.

[7] A. A. Fingelkurts, A. A. Fingelkurts, and S. Kähkönen, "Functional connectivity in the brain—is it an elusive concept?" *Neuroscience & Biobehavioral Reviews*, vol. 28, no. 8, pp. 827–836, 2005.

[8] B. Lei, Y. Liang, J. Xie, Y. Wu, E. Liang, Y. Liu, P. Yang, T. Wang, C. Liu, J. Du *et al.*, "Hybrid federated learning with brain-region attention network for multi-center alzheimer's disease detection," *Pattern Recognition*, p. 110423, 2024.

[9] 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.

[10] H. Huang, X. Hu, Y. Zhao, M. Makkie, Q. Dong, S. Zhao, L. Guo, and T. Liu, "Modeling task fmri data via deep convolutional autoencoder," *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1551–1561, 2017.

[11] Z.-A. Huang, Z. Zhu, C. H. Yau, and K. C. Tan, "Identifying autism spectrum disorder from resting-state fmri using deep belief network," *IEEE Transactions on neural networks and learning systems*, vol. 32, no. 7, pp. 2847–2861, 2020.

[12] K. Zhao, B. Duka, H. Xie, D. J. Oathes, V. Calhoun, and Y. Zhang, "A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in adhd," *Neuroimage*, vol. 246, p. 118774, 2022.

[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. Yang, C. Ye, X. Guo, T. Wu, Y. Xiang, and T. Ma, "Mapping multimodal brain connectome for brain disorder diagnosis via cross-modal mutual learning," *IEEE Transactions on Medical Imaging*, 2023.

[15] Z. Qiu, P. Yang, C. Xiao, S. Wang, X. Xiao, J. Qin, C.-M. Liu, T. Wang, and B. Lei, "3d multimodal fusion network with disease-induced joint learning for early alzheimer's disease diagnosis," *IEEE Transactions on Medical Imaging*, 2024.

[16] X. Kan, W. Dai, H. Cui, Z. Zhang, Y. Guo, and C. Yang, "Brain network transformer," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25 586–25 599, 2022.

[17] Q. Zhu, H. Wang, B. Xu, Z. Zhang, W. Shao, and D. Zhang, "Multimodal triplet attention network for brain disease diagnosis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3884–3894, 2022.

[18] G. Wen, P. Cao, L. Liu, J. Yang, X. Zhang, F. Wang, and O. R. Zaiane, "Graph self-supervised learning with application to brain networks analysis," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[19] W. Jung, D.-W. Heo, E. Jeon, J. Lee, and H.-I. Suk, "Inter-regional high-level relation learning from functional connectivity via self-supervision," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*. Springer, 2021, pp. 284–293.

[20] A. Segal, L. Parkes, K. Aquino, S. M. Kia, T. Wolfers, B. Franke, M. Hoogman, C. F. Beckmann, L. T. Westlye, O. A. Andreassen *et al.*, "Regional, circuit and network heterogeneity of brain abnormalities in psychiatric disorders," *Nature Neuroscience*, vol. 26, no. 9, pp. 1613–1629, 2023.

[21] Y. Zhang, L. Zhan, W. Cai, P. Thompson, and H. Huang, "Integrating heterogeneous brain networks for predicting brain disease conditions," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22*. Springer, 2019, pp. 214–222.

[22] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.

[23] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," *arXiv preprint arXiv:2003.04297*, 2020.

[24] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[25] A. Conneau and G. Lample, "Cross-lingual language model pretraining," *Advances in neural information processing systems*, vol. 32, 2019.

[26] S. Zhang and D. Metaxas, "On the challenges and perspectives of foundation models for medical image analysis," *arXiv preprint arXiv:2306.05705*, 2023.

[27] J. Hu, Y. Huang, N. Wang, and S. Dong, "Brainnpt: Pre-training of transformer networks for brain network classification," *arXiv preprint arXiv:2305.01666*, 2023.

[28] B. Lei, Y. Zhu, E. Liang, P. Yang, S. Chen, H. Hu, H. Xie, Z. Wei, F. Hao, X. Song *et al.*, "Federated domain adaptation via transformer for multi-site alzheimer's disease diagnosis," *IEEE Transactions on Medical Imaging*, 2023.

[29] B. Lei, Y. Zhu, S. Yu, H. Hu, Y. Xu, G. Yue, T. Wang, C. Zhao, S. Chen, P. Yang *et al.*, "Multi-scale enhanced graph convolutional network for mild cognitive impairment detection," *Pattern Recognition*, vol. 134, p. 109106, 2023.

[30] B. Lei, E. Liang, M. Yang, P. Yang, F. Zhou, E.-L. Tan, Y. Lei, C.-M. Liu, T. Wang, X. Xiao *et al.*, "Predicting clinical scores for alzheimer's disease based on joint and deep learning," *Expert Systems with Applications*, vol. 187, p. 115966, 2022.

[31] B. Jie, M. Liu, C. Lian, F. Shi, and D. Shen, "Designing weighted correlation kernels in convolutional neural networks for functional connectivity based brain disease diagnosis," *Medical image analysis*, vol. 63, p. 101709, 2020.

[32] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, 2018.

[33] Y. Wang, Y. Yang, X. Guo, C. Ye, N. Gao, Y. Fang, and H. T. Ma, "A novel multimodal mri analysis for alzheimer's disease based on convolutional neural network," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 754–757.

[34] N. S. Dsouza, M. B. Nebel, D. Crocetti, J. Robinson, S. Mostofsky, and A. Venkataraman, "M-gcn: A multimodal graph convolutional network to integrate functional and structural connectomics data to predict multidimensional phenotypic characterizations," in *Medical Imaging with Deep Learning*. PMLR, 2021, pp. 119–130.

[35] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert, "Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease," *Medical image analysis*, vol. 48, pp. 117–130, 2018.

[36] X. Song, F. Zhou, A. F. Frangi, J. Cao, X. Xiao, Y. Lei, T. Wang, and B. Lei, "Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction," *Medical Image Analysis*, vol. 69, p. 101947, 2021.

[37] Y. Yang, C. Ye, and T. Ma, "A deep connectome learning network using graph convolution for connectome-disease association study," *Neural Networks*, 2023.

[38] A. Kazi, S. Shekarforoush, S. Arvind Krishna, H. Burwinkel, G. Vivar, K. Kortüm, S.-A. Ahmadi, S. Albarqouni, and N. Navab, "Inceptiongcn: receptive field aware graph convolutional network for disease prediction," in *Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26*. Springer, 2019, pp. 73–85.

[39] C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, and T.-Y. Liu, "Do transformers really perform badly for graph representation?" *Advances in Neural Information Processing Systems*, vol. 34, pp. 28 877–28 888, 2021.

[40] D. Kreuzer, D. Beaini, W. Hamilton, V. Létourneau, and P. Tossou, "Rethinking graph transformers with spectral attention," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21 618–21 629, 2021.

[41] X. Wang, X. Zhang, G. Wang, J. He, Z. Li, W. Zhu, Y. Guo, Q. Dou, X. Li, D. Wang *et al.*, "Openmedlab: An open-source platform for multi-modality foundation models in medicine," *arXiv preprint arXiv:2402.18028*, 2024.

[42] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.

[43] Y. Yang, G. Xutao, C. Ye, Y. Xiang, and T. Ma, "Regularizing brain age prediction via gated knowledge distillation," in *International Conference on Medical Imaging with Deep Learning*. PMLR, 2022, pp. 1430–1443.

[44] J. Ortega Caro, A. H. Oliveira Fonseca, C. Averill, S. A. Rizvi, M. Rosati, J. L. Cross, P. Mittal, E. Zappala, D. Levine, R. M. Dhodapkar *et al.*, "Brainlm: A foundation model for brain activity recordings," *bioRxiv*, pp. 2023–09, 2023.

[45] A. Thomas, C. Ré, and R. Poldrack, "Self-supervised learning of brain dynamics from broad neuroimaging data," *Advances in neural information processing systems*, vol. 35, pp. 21 255–21 269, 2022.

[46] S. Sihag, G. Mateos, and A. Ribeiro, "Towards a foundation model for brain age prediction using covariance neural networks," *arXiv preprint arXiv:2402.07684*, 2024.

[47] S. Azizi, B. Mustafa, F. Ryan, Z. Beaver, J. Freyberg, J. Deaton, A. Loh, A. Karthikesalingam, S. Kornblith, T. Chen *et al.*, "Big self-supervised models advance medical image classification," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 3478–3488.

[48] Z. Wang, C. Liu, S. Zhang, and Q. Dou, "Foundation model for endoscopy video analysis via large-scale self-supervised pre-train," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 101–111.

[49] X. Zhang, C. Wu, Y. Zhang, W. Xie, and Y. Wang, "Knowledge-enhanced visual-language pre-training on chest radiology images," *Nature Communications*, vol. 14, no. 1, p. 4542, 2023.

[50] Y. Zhou, M. A. Chia, S. K. Wagner, M. S. Ayhan, D. J. Williamson, R. R. Struyven, T. Liu, M. Xu, M. G. Lozano, P. Woodward-Court *et al.*, "A foundation model for generalizable disease detection from retinal images," *Nature*, vol. 622, no. 7981, pp. 156–163, 2023.

[51] Z. Huang, F. Bianchi, M. Yuksekgonul, T. J. Montine, and J. Zou, "A visual–language foundation model for pathology image analysis using medical twitter," *Nature medicine*, vol. 29, no. 9, pp. 2307–2316, 2023.

[52] D. Chen, Y. Lin, W. Li, P. Li, J. Zhou, and X. Sun, "Measuring and relieving the over-smoothing problem for graph neural networks from the topological view," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 04, 2020, pp. 3438–3445.

[53] M. Liu, H. Gao, and S. Ji, "Towards deeper graph neural networks," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 338–348.

[54] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[55] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray *et al.*, "Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age," *Plos med*, vol. 12, no. 3, p. e1001779, 2015.

[56] L. M. Alexander, J. Escalera, L. Ai, C. Andreotti, K. Febre, A. Mangone, N. Vega-Potler, N. Langer, A. Alexander, M. Kovacs *et al.*, "An open resource for transdiagnostic research in pediatric mental health and learning disorders," *Scientific data*, vol. 4, no. 1, pp. 1–26, 2017.

[57] A. J. Holmes, M. O. Hollinshead, T. M. O'keefe, V. I. Petrov, G. R. Fariello, L. L. Wald, B. Fischl, B. R. Rosen, R. W. Mair, J. L. Roffman *et al.*, "Brain genomics superstruct project initial data release with structural, functional, and behavioral measures," *Scientific data*, vol. 2, no. 1, pp. 1–16, 2015.

[58] X.-N. Zuo, J. S. Anderson, P. Bellec, R. M. Birn, B. B. Biswal, J. Blautzik, J. Breitner, R. L. Buckner, V. D. Calhoun, F. X. Castellanos *et al.*, "An open science resource for establishing reliability and reproducibility in functional connectomics," *Scientific data*, vol. 1, no. 1, pp. 1–13, 2014.

[59] R. H. Tobe, A. MacKay-Brandt, R. Lim, M. Kramer, M. M. Breland, L. Tu, Y. Tian, K. D. Trautman, C. Hu, R. Sangoi *et al.*, "A longitudinal resource for studying connectome development and its psychiatric associations during childhood," *Scientific Data*, vol. 9, no. 1, p. 300, 2022.

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

[61] B. Sinclair, N. K. Hansell, G. A. Blokland, N. G. Martin, P. M. Thompson, M. Breakspear, G. I. de Zubicaray, M. J. Wright, and K. L. McMahon, "Heritability of the network architecture of intrinsic brain functional connectivity," *Neuroimage*, vol. 121, pp. 243–252, 2015.

[62] D. Wei, K. Zhuang, L. Ai, Q. Chen, W. Yang, W. Liu, K. Wang, J. Sun, and J. Qiu, "Structural and functional brain scans from the cross-sectional southwest university adult lifespan dataset," *Scientific data*, vol. 5, no. 1, pp. 1–10, 2018.

[63] X.-N. Zuo and C. Consortium, "Chinese Color Nest Project (CCNP)," Feb. 2023.

[64] J. R. Taylor, N. Williams, R. Cusack, T. Auer, M. A. Shafto, M. Dixon, L. K. Tyler, R. N. Henson *et al.*, "The cambridge centre for ageing and neuroscience (cam-can) data repository: Structural and functional mri, meg, and cognitive data from a cross-sectional adult lifespan sample," *neuroimage*, vol. 144, pp. 262–269, 2017.

[65] L. T. Strike, N. K. Hansell, K.-H. Chuang, J. L. Miller, G. I. de Zubicaray, P. M. Thompson, K. L. McMahon, and M. J. Wright, "The queensland twin adolescent brain project, a longitudinal study of adolescent brain development," *Scientific Data*, vol. 10, no. 1, p. 195, 2023.

[66] R. N. Spreng, R. Setton, U. Alter, B. N. Cassidy, B. Darboh, E. DuPre, K. Kantarovich, A. W. Lockrow, L. Mwilambwe-Tshilobo, W.-M. Luh *et al.*, "Neurocognitive aging data release with behavioral, structural and multi-echo functional mri measures," *Scientific Data*, vol. 9, no. 1, p. 119, 2022.

[67] P. Gao, H.-M. Dong, Y.-S. Wang, C.-S. Yu, and X.-N. Zuo, "Imaging Chinese Young Brains (I See Your Brain)," Aug. 2021.

[68] J. W. Kable, M. K. Caulfield, M. Falcone, M. McConnell, L. Bernardo, T. Parthasarathi, N. Cooper, R. Ashare, J. Audrain-McGovern, R. Hornik *et al.*, "No effect of commercial cognitive training on brain activity, choice behavior, or cognitive performance," *Journal of Neuroscience*, vol. 37, no. 31, pp. 7390–7402, 2017.

[69] D. Kliemann, R. Adolphs, T. Armstrong, P. Galdi, D. A. Kahn, T. Rusch, A. Z. Enkavi, D. Liang, S. Lograsso, W. Zhu *et al.*, "Caltech conte center, a multimodal data resource for exploring social cognition and decision-making," *Scientific Data*, vol. 9, no. 1, p. 138, 2022.

[70] C. Racey, C. Kampoureli, O. Bowen-Hill, M. Bauer, I. Simpson, C. Rae, M. Del Rio, J. Simner, and J. Ward, "An open science mri database of over 100 synaesthetic brains and accompanying deep phenotypic information," *Scientific Data*, vol. 10, no. 1, p. 766, 2023.

[71] X. Chen, B. Lu, H.-X. Li, X.-Y. Li, Y.-W. Wang, F. X. Castellanos, L.-P. Cao, N.-X. Chen, W. Chen, Y.-Q. Cheng *et al.*, "The direct consortium and the rest-meta-mdd project: towards neuroimaging biomarkers of major depressive disorder," *Psychoradiology*, vol. 2, no. 1, pp. 32–42, 2022.

[72] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. G. Vlassenko *et al.*, "Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease," *MedRxiv*, pp. 2019–12, 2019.

[73] C. R. Jack Jr, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward *et al.*, "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 27, no. 4, pp. 685–691, 2008.

[74] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, S. Chowdhury *et al.*, "The parkinson progression marker initiative (ppmi)," *Progress in neurobiology*, vol. 95, no. 4, pp. 629–635, 2011.

[75] R. A. Poldrack, E. Congdon, W. Triplett, K. Gorgolewski, K. Karlsgodt, J. Mumford, F. Sabb, N. Freimer, E. London, T. Cannon *et al.*, "A phenome-wide examination of neural and cognitive function," *Scientific data*, vol. 3, no. 1, pp. 1–12, 2016.

[76] Y. Yang, C. Ye, J. Sun, L. Liang, H. Lv, L. Gao, J. Fang, T. Ma, and T. Wu, "Alteration of brain structural connectivity in progression of parkinson's disease: a connectome-wide network analysis," *NeuroImage: Clinical*, vol. 31, p. 102715, 2021.

[77] L. Wang, K. I. Alpert, V. D. Calhoun, D. J. Cobia, D. B. Keator, M. D. King, A. Kogan, D. Landis, M. Tallis, M. D. Turner *et al.*, "Schizconnect: Mediating neuroimaging databases on schizophrenia and related disorders for large-scale integration," *Neuroimage*, vol. 124, pp. 1155–1167, 2016.

[78] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar *et al.*, "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in neural information processing systems*, vol. 33, pp. 21 271–21 284, 2020.

[79] 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.

[80] J. Jiang, Y. Wei, Y. Feng, J. Cao, and Y. Gao, "Dynamic hypergraph neural networks." in *IJCAI*, 2019, pp. 2635–2641.

[81] Y. Yang, X. Guo, C. Ye, Y. Xiang, and T. Ma, "Creg-kd: Model refinement via confidence regularized knowledge distillation for brain imaging," *Medical Image Analysis*, vol. 89, p. 102916, 2023.

[82] U. Mahmood, Z. Fu, S. Ghosh, V. Calhoun, and S. Plis, "Through the looking glass: Deep interpretable dynamic directed connectivity in resting fmri," *NeuroImage*, vol. 264, p. 119737, 2022.

[83] A. Mitra and M. E. Raichle, "Principles of cross-network communication in human resting state fmri," *Scandinavian Journal of Psychology*, vol. 59, no. 1, pp. 83–90, 2018.

[84] A. I. Luppi, P. A. Mediano, F. E. Rosas, N. Holland, T. D. Fryer, J. T. O'Brien, J. B. Rowe, D. K. Menon, D. Bor, and E. A. Stamatakis, "A synergistic core for human brain evolution and cognition," *Nature Neuroscience*, vol. 25, no. 6, pp. 771–782, 2022.

[85] T. Ito, K. R. Kulkarni, D. H. Schultz, R. D. Mill, R. H. Chen, L. I. Solomyak, and M. W. Cole, "Cognitive task information is transferred between brain regions via resting-state network topology," *Nature communications*, vol. 8, no. 1, p. 1027, 2017.

[86] Z. Shehzad, C. Kelly, P. T. Reiss, R. C. Craddock, J. W. Emerson, K. McMahon, D. A. Copland, F. X. Castellanos, and M. P. Milham, "A multivariate distance-based analytic framework for connectome-wide association studies," *Neuroimage*, vol. 93, pp. 74–94, 2014.

[87] M. Wei, J. Qin, R. Yan, K. Bi, C. Liu, Z. Yao, and Q. Lu, "Association of resting-state network dysfunction with their dynamics of inter-network interactions in depression," *Journal of affective disorders*, vol. 174, pp. 527–534, 2015.

[88] J. Caspers, C. Rubbert, S. B. Eickhoff, F. Hoffstaedter, M. Südmeyer, C. J. Hartmann, B. Sigl, N. Teichert, J. Aissa, B. Turowski *et al.*, "Within-and across-network alterations of the sensorimotor network in parkinson's disease," *Neuroradiology*, vol. 63, no. 12, pp. 2073–2085, 2021.

[89] Z. Qi, Y. An, M. Zhang, H.-J. Li, and J. Lu, "Altered cerebro-cerebellar limbic network in ad spectrum: a resting-state fmri study," *Frontiers in Neural Circuits*, vol. 13, p. 72, 2019.

[90] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685*, 2021.

[91] S.-Y. Liu, C.-Y. Wang, H. Yin, P. Molchanov, Y.-C. F. Wang, K.-T. Cheng, and M.-H. Chen, "Dora: Weight-decomposed low-rank adaptation," *arXiv preprint arXiv:2402.09353*, 2024.