Hypercomplex-valued Graph Neural Network: Towards Deep Intersection Coupling of Multi-modal Brain Networks

Yanwu Yang, *Member, IEEE*, Guoqing Cai, Chenfei Ye, Kunru Song, Jintao Zhang, Yang Xiang, Ting Ma *Member, IEEE*

***Abstract*—The study of multi-modal brain networks has greatly improved the efficiency of disease diagnosis and offers insights into understanding the heteromodal relationships between brain network organization and behavioral phenotypes. The integration of data with multiple modalities facilitates characterizing the interplay of anatomical, functional, and physiological brain alterations or development. Graph Neural Networks (GNNs) have recently gained popularity in the analysis and fusion of multi-modal graph-structured brain networks. However, there is still a great challenge to learn from other modalities for complementary representations effectively, due to the sophisticated and heterogeneous inter-modal dependencies. Moreover, most existing studies often focus on specific modalities (e.g., only two modalities or only fMRI and DTI), limiting their scalability to other types of modalities. To address these limitations, we propose a HyperComplex-valued Graph Neural Network (HC-GNN) that models multi-modal networks as hypercomplex tensor graphs. We investigate inter-modal dependencies by modelling HC-GNN as a dynamic spatial graph, where the attentively learned inter-modal associations are represented as the adjacency matrix. We carry out statistical analysis on the saliency maps for associating disease biomarkers. Extensive experiments on three datasets demonstrate the superior classification performance of our method and its strong scalability to various types of modalities. Our work presents a powerful paradigm for the study of multi-modal brain networks.**

***Index Terms*—Multi-modal learning, graph neural network (GNN), neuroimage, hypercomplex-GNN**

This work is 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 (62106113, 62276081, and 62106115), Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (NO: ZYYCXTD-C-202004), The National Key Research and Development Program of China (2021YFC2501202), and the Major Key Project of PCL. Corresponding author: Ting Ma.

Yanwu Yang is 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, 18b952052@stu.hit.edu.cn)

Guoqing Cai is with the School of Electronics and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, China. (e-mail: cgq982023711@163.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, Guangdong Provincial Key Laboratory of Aerospace Communication and Networking Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China, and International Research Institute for Artificial Intelligence, Harbin Institute of Technology (Shenzhen), Shenzhen, China. (e-mail: tma@hit.edu.cn)

Chenfei Ye is with the International Research Institute for Artificial Intelligence, Harbin Institute of Technology at Shenzhen, Shenzhen, China. (e-mail: Chenfei.ye@foxmail.com)

Yang Xiang is with the Peng Cheng Laboratory, Shenzhen, Guangdong, China. (e-mail: xiangy@pcl.ac.cn)

Kunru Song and Jintao Zhang are with the State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China. (e-mail: songkunru@mail.bnu.edu.cn, zhangjintao@bnu.edu.cn)

I. Introduction

THE pursuit of decoding brain network has provided us with insights into the neural mechanisms of the brain and the intricate relationships between individual brain dysfunctions and behavioral phenotypes [1]–[4]. Recent investigations have made significant progress in examining and identifying connectome patterns in brain disorders such as autism spectrum disorder (ASD) and Alzheimer's disease (AD), as well as behavioral phenotypes including full-scale intelligence quotient (FSIQ), sex and age differences [5]–[9]. Furthermore, recent studies that conceptualize the brain as a network have offered a more comprehensive perspective in understanding the relationship between abnormal brain activity and dysfunction. In this line, Graph Neural Networks (GNNs) have been extensively utilized for connectome-wide association studies (CWAS) and have emerged as state-of-the-art tools for exploring the topological and spatial organization of brain networks.

Recently, the integration of multi-modal networks has emerged as a significant area of research. Brain networks encompass functional networks derived from functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), as well as structural networks constructed from diffusion tensor imaging (DTI) or diffusion spectrum imaging (DSI) [10], [11]. Exploring the relationships between different modalities enables us to capture important complementary information and enhance representations. Previous studies suggest that multiple types of brain connectivity might be mediated by each other and link multi-modal connectome for mediating [12], [13]. Additionally, subtle abnormalities or misleading representations within a single modality can be augmented by incorporating another modality [14]. Moreover, multi-modal studies allow us to explore brain states by examining neuron activation and connections in vivo, providing a more comprehensive understanding with distinct biomarkers [15], [16]. Accordingly, it is natural and well-justified to combine multi-modal information for brain network studies.

One of the challenges in multi-modal brain network studies is that the multiple modalities usually have sophisticated and

heterogeneous structures. Simply applying feature embedding to the entire feature set is inefficient for capturing meaningful representations across all modalities. Additionally, it is nontrivial to effectively learn from each other modality to enrich representations. Furthermore, medical data are typically high-dimensional, but the available data samples are limited. Incorporating a large number of features for multi-modal learning would introduce unwanted noise and potentially lead to overfitting that decreases performance. Machine learning approaches implement feature learning methods for capturing meaningful features, however, are limited in their performance. Lastly, most existing models are designed for specific modalities and lack scalability and generalizability to other types of modalities.

A straightforward and natural way of exploring multimodal data is to model multimodal data as multi-dimensional arrays, known as tensors. Tensors have proven to be powerful and promising tools for uncovering complex underlying structures in data, allowing for the separation of common and independent components [17]. Most tensor-based methods employ tensor decomposition techniques, such as Canonical Polyadic (CP) decomposition and Tucker Decomposition (TD), to identify key features and improve learning efficiency [18]. However, most of these approaches have limitations in terms of lossy transformations and may overlook important intrinsic complementary information, resulting in suboptimal results.

In this regard, we propose a HyperComplex-valued Graph Neural Network (HC-GNN) for multi-modal brain network learning, which enables to enrich features and the deep interactions of representations through cross-embedding and cross-aggregation in an end-to-end manner. Specifically, we construct tensor graphs by formulating multimodal node features as tensors and further implement hypercomplex-valued neural network (HNN) [19], [20] to encode tensor representations. The approach involves constructing composite tensors to learn latent multimodal representations and generate high-order generalizations of matrices to capture the consistency among different modalities. Moreover, the deep intersection behind HNN allows for richer representations by expressing neuron outputs with multiple indices.

In addition, in this study, we propose to implement the dynamic mechanism to capture the inter-modal feature dependencies to relate attentive multi-modal dependency as the adjacency matrix. The approach adaptively characterizes the intrinsic associations between different modalities and addresses the issue of unknown node relationships in multi-modal graphs. Moreover, most existing multi-modal networks overlook the importance of neuronal synchrony, which involves information gating and dynamic binding of inter-modal representations. To address this, we incorporate a gating mechanism to control information flow and dynamically bind inter-modal associations in an attentive manner. Extensive experiments on three real-world datasets demonstrate that our proposed method is a superior tool for multi-modal brain network representation learning, which outperforms other baselines in terms of prediction performance, generalizability and scalability.

The rest of our paper is structured as follows. We would like to review related methods in terms of connectome study, multi-modal neuroimaging, and hypercomplex-valued approaches in Section II. The details of the proposed model are introduced in Section III. Section IV describes the experiment settings. The results are provided and discussed in Section V and VI. Section VII draws the conclusions of the work.

II. Related works

A. Brain Connectome Study

Decoding the brain network provides valuable insights into the relationship between brain network organization and behavioral characteristics. Connectome-wide association studies (CWAS) have been successful in identifying dysconnectivity patterns in brain disorders. Statistical tools such as Network-based statistical (NBS) analysis [21] and multivariate distance matrix regression (MDMR) [22] are commonly used for quantifying connectome reorganization across the entire brain network. Machine learning models, such as SVM and random forest, are then employed for classification using the identified key biomarkers [23], [24]. However, these approaches have limitations in terms of manually extracting features and often yield limited performance.

Deep learning approaches have prompted a shift in connectome studies towards subject-level predictions with promising accuracy. These approaches include Multi-layer Perceptions (MLP) [25], Convolutional Neural Networks (CNN) [26], [27], and Graph Neural Networks (GNN) [28]–[30]. In particular, BrainNetCNN treats brain connectome networks as grid-like data and proposes a convolutional neural network to capture topological locality [26]. [31] suggests using Deep Belief Networks (DBNs) on functional connectivity to distinguish autism spectrum disorder. Graph-based methods have also been proposed to investigate graph isomorphism and structural topology properties [32], [33]. Although the connectome incorporating the strength of connections in their edges assumes a graph-like organization, the connections are limited to depict sophisticated and non-linear brain structure relationships [34]. To address the issue of unclear graph structure, several strategies have been proposed to avoid potential aggregation of misleading contextual information [35], [36]. DGCNN dynamically gathers nearby information to recover topology for learning [37]. [34] suggests learning the graph structure based on the small-world model. [38] aggregates the top-k nearest nodes to update node features. In this study, we employ the dynamic mechanism to model attentive inter-node dependencies.

B. Multi-modal graph learning

Multi-modal neuroimaging technologies have provided unprecedented opportunities for disease diagnosis and connectome-behavior association studies. Deep learning methods are feasible to capture more meaningful features and contribute to a better understanding of brain network. For example, [27] proposes a two-layer convolution approach that simultaneously processes fMRI and DTI data. M-GCN introduces regularization of convolution on functional connectivity using structural graph Laplacian [39]. [40] implements a

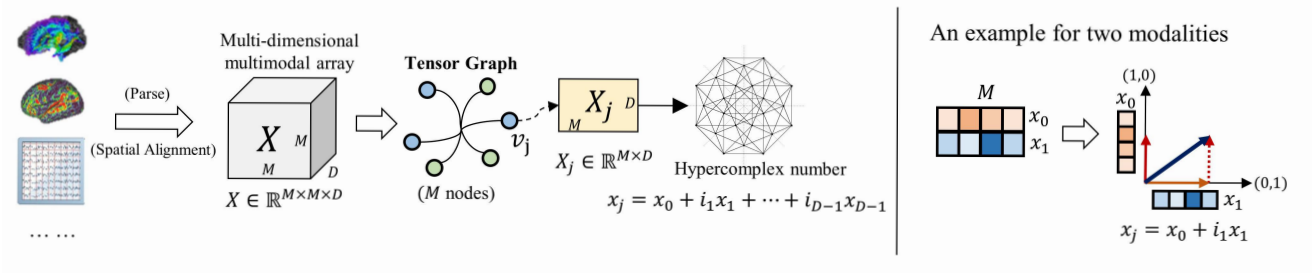

Fig. 1. Tensor graph construction. First, we perform spatial alignment to map multi-modal brain networks into the same space. Next, we formulate the concatenated multi-modal multi-dimensional array as a tensor graph. In this graph, there are M nodes, with each node represented as a tensor.

fusion network to integrate functional and structural magnetic resonance images.

Furthermore, well-established graph-based models can effectively capture multi-modal representations. For instance, [41] constructs a graph with variational edges that incorporate multimodal features. [42] applies graph hashing learning to preserve the original semantic spatial relationships. The use of hypergraphs provides a strategy to capture dependencies among nodes by constructing hyperedges based on each modality. Examples include Hypergraph Neural Networks (HNN) [43] and Dynamic Hypergraph Neural Networks (DHGNN) [44]. Additionally, [45] proposes a graph parser and matcher to capture the correspondence between nodes from different modalities. In another approach, MMP-GCN proposes a penalty term to fuse one modality using multi-center and multi-channel mechanisms [46]. Triplet Attention Network introduces a triplet network with self-attention to map high-order multi-modal representations [47]. However, many of these studies either overlook important complementary information between modalities or lack generalization capabilities to handle data with more modalities.

C. Hypercomplex and Complex-valued neural networks

Recently, there has been increasing attention on Hypercomplex-valued neural networks (HNNs) and Complex-valued neural networks (CVNNs) in signal processing, as they have shown the ability to achieve competitive performance while reducing parameters [48], [49]. HNNs offer a new perspective for enhancing representational capacity by utilizing hypercomplex numbers instead of the commonly used real-valued numbers. By leveraging the interactions among the real and imaginary units, HNNs can capture internal relationships in inputs and preserve preexisting correlations [20], [50], [51]. Moreover, studies have demonstrated that hypercomplex representations provide better generalization performance [52], can correlate information from multiple fields [53], and improve gradient regularization [54].

In recent years, there has been a growing body of research exploring the potential use of hypercomplex neural networks in various domains such as image processing, natural language processing, and signal processing [55], [56]. For instance, [57] applied complex-valued neural networks for object detection in images, with depth sensors represented as the imaginary part. [20] introduced a family of parameterized HNNs and their applications in multi-dimensional images. [58] developed a quaternion convolutional neural network for automatic speech recognition, and [59] utilized hypercomplex context for scene graph generation. However, there have been few studies investigating the potential use of hypercomplex neural networks in multi-modal graph learning. In this work, we address multi-modal representation learning by formulating multi-modal data as hypercomplex tensors, which is extended from our previous study [60]. We represent multiple modalities as real and multi-imaginary parts to facilitate dynamic interactions.

III. Method

A. Problem Definition

Brain Network Graph: The derived brain networks from neuroimages or EEG are symmetric postitive semidefine matrices $X \in \mathbb{R}^{M\times M}$, which are derived by mapping the input into a template with M regions. Each element x_{ij} of X denotes a co-variance or connectivity strength between the region i and j. Traditionally, the brain network is formulated as a graph $G = (\mathcal{V}, \mathcal{E})$, where the sets $\mathcal{V}$ and $\mathcal{E}$ denote the node and edges respectively. For each vertex $v_i \in \mathbb{R}^m$, the node feature is built by the i-th column or row in the derived matrices. The edges $\mathcal{E} \in \mathbb{R}^{m\times m}$ are represented by the matrices X directly.

Dynamic Graph: Unlike most existing graph studies with explicit and clear graph structures, the brain network graphs remain unknown, where the derived matrices fail to model sophisticated relationships of the brain regions. The dynamic mechanism has been widely used for modeling brain graphs [34], [38], [61]. In detail, for a graph $G = (\mathcal{V}, \mathcal{E})$, all the node features $H \in \mathbb{R}^{M\times M}$ and adjacency matrix $A \in \mathbb{R}^{M\times M}$ are obtained by learning mappings from the input: $f_{\mathcal{V}} : \mathcal{X} \to \mathcal{V}, f_{\mathcal{E}} : \mathcal{X} \to \mathcal{E}$. The mapped graphs are denoted as dynamic graphs.

Multi-modal brain graph learning: In this study, we leverage various types of multimodal brain networks to explore connectome-phenotype associations. The pre-processed brain networks are mapped into the same spatial template. Accordingly, all types of brain networks are in the same dimension, $X_0, X_1, ..., X_{D-1} \in \mathbb{R}^{M\times M}$. An input is expressed as $\hat{G} = \{G_0, G_1, ..., G_{D-1}\}$, where G_d denote the d-th modal graph. Formally, given a set of graphs $\mathcal{G} = \{\hat{G}_1, \hat{G}_2, ..., \hat{G}_N\}$ with a few labeled graph instances, the goal of the study is to distinguish the disease or the phenotype as a graph multi-modal classification task with a mapping $f_{\mathcal{G}} : \mathcal{G} \to \mathcal{Y}$.

B. Hypercomplex Tensor Graph Construction

As shown in Figure 1, our first step is to parse and map different types of brain networks into a unified space through spatial alignment. The specific details of spatial alignment will be introduced in the pre-processing experiments section. Subsequently, the pre-processed arrays are transformed into a tensor graph for each participant. This graph consists of M nodes, with each node represented as an array $X_j \in \mathbb{R}^{M\times D}$. Here, M corresponds to the number of regions of interest (ROIs) in the brain network, and D represents the number of modalities incorporated.

Node. The features of each modality are represented in real or one imaginary unit. We provide a common example in Figure 1 involving two modalities. In the case, one modality is represented as the real part, and the other is represented as the imaginary part. For more modalities, multiple bases (e.g., $(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)$ for four modalities) are utilized.

Edge. We apply the dynamic strategy to model edges dynamically by a mapping as: $f_{\mathcal{E}} : \{\mathbf{X}_0, \mathbf{X}_1, ..., \mathbf{X}_{D-1}\} \rightarrow \mathcal{E}$. The mapping is obtained by learning an embedding though hypercomplex convolution.

C. Preliminaries of complex-valued neural networks

In this section, we would like to firstly introduce the complex-valued operations in neural networks, and extend it into the hypercomplex.

*1) **Complex convolution.**:* For a complex filter matrix $\mathbf{W} = \mathbf{W}_r + i\mathbf{W}_i$, the convolution on an input $\mathbf{x} = \mathbf{x}_r + i\mathbf{x}_i$ is expressed as:

$$\mathbf{W} * \mathbf{x} = (\mathbf{W}_r\mathbf{x}_r - \mathbf{W}_i\mathbf{x}_i) + i(\mathbf{W}_i\mathbf{x}_r + \mathbf{W}_r\mathbf{x}_i) \quad (1)$$

The $\mathbf{W}_r$ and $\mathbf{x}_r$ denote the real parts and $\mathbf{W}_i$, $\mathbf{x}_i$ are imaginary parts, and $i = \sqrt{-1}$. For multi-modal learning, this function can be considered as a convolution by a filter matrix $\mathbf{W}_r$, leading to $\mathbf{W}_r\mathbf{x_r} + i\mathbf{W}_r\mathbf{x_i}$, and calibrated with another field $\mathbf{W}_i$. In addition, matrix multiply operation is in the same way by reference to Eq. (1).

*2) **Complex-valued activation function.**:* Activation functions play essential roles for locating meaningful nonlinear representations. Previous studies have proposed activation functions for complex-valued networks including $\mathbb{C}$ReLU, ModReLU, zReLU and TReLU [19], [62]:

$$\mathbb{C}ReLU(\mathbf{x}) = ReLU(\mathbf{x}_r) + iReLU(\mathbf{x}_i) \quad (2)$$

$\mathbb{C}ReLU$ enables non-linear activation on the real and imaginary parts respectively.

$$modReLU(\mathbf{x}) = (|\mathbf{x}| + b)\frac{\mathbf{x}}{|\mathbf{x}|}, \text{if } |\mathbf{x}| + b > 0, \text{else } 0. \quad (3)$$

$modReLU$ maintains the phase after the activation and leverage a trainable bias parameter b.

$$zReLU(\mathbf{x}) = z, \text{ if } \phi(\mathbf{x}) \in [0, \frac{\pi}{2}] \quad (4)$$

$zReLU$ filters activation maps when the phase is in the range of $[0, \frac{\pi}{2}]$.

*3) **Complex-valued Readout**:* Given a complex-valued feature vector, a readout function is to convert the complex numbers into real by $\text{Readout}(\mathbf{x}) = \sqrt{\mathbf{x}_r^2 + \mathbf{x}_i^2}, \mathbf{x} \in \mathbb{C}$.

D. From complex to hypercomplex

The complex-valued operations are only available for complex numbers. Since we model the modalities with the real and multiple imaginary units, the complex-valued neural network is only able to model two-modality data. In this regard, hypercomplex neural networks are feasible to generalize to more modalities. For a basis of $\{1, i_1, i_2, ..., i_{D-1}\}$, an input with D modalities is formulated as:

$$\mathbf{x} = \mathbf{x}_0 + i_1\mathbf{x}_1 + i_2\mathbf{x}_2 + ... + i_{D-1}\mathbf{x}_{D-1} \quad (5)$$

The corresponding complex operations are reformulated into the hypercomplex. For example, the addition of two hypercomplex numbers $\mathbf{x} = \mathbf{x}_0 + i_1\mathbf{x}_1 + i_2\mathbf{x}_2 + ... + i_{D-1}\mathbf{x}_{D-1}$ and $\mathbf{y} = \mathbf{y}_0 + i_1\mathbf{y}_1 + i_2\mathbf{y}_2 + ... + i_{D-1}\mathbf{y}_{D-1}$ is defined in a component-wise manner as:

$$\mathbf{x}+\mathbf{y} = (\mathbf{x}_0+\mathbf{y}_0)+i_1(\mathbf{x}_1+\mathbf{y}_1)+...+i_{D-1}(\mathbf{x}_{D-1}+\mathbf{y}_{D-1}) \quad (6)$$

The product between $\mathbf{x}$ and $\mathbf{y}$, denoted by the juxtaposition of $\mathbf{x}$ and $\mathbf{y}$, is given by:

$$\begin{aligned}\mathbf{x}\otimes\mathbf{y} = &(\mathbf{x}_0\mathbf{y}_0 + \sum_{u,v=1}^{D-1}\mathbf{x}_u\mathbf{y}_v a_{uv,0}) \\ &+ i_1(\mathbf{x}_0\mathbf{y}_1 + \mathbf{x}_1\mathbf{y}_0 + \sum_{u,v=1}^{D-1}\mathbf{x}_u\mathbf{y}_v a_{uv,1}) + ... \\ &+ i_{D-1}(\mathbf{x}_0\mathbf{y}_{D-1} + \mathbf{x}_{D-1}\mathbf{y}_0 + \sum_{u,v=1}^{D-1}\mathbf{x}_u\mathbf{y}_v a_{uv,D-1})\end{aligned} \quad (7)$$

where the product between i_u and i_v is the hypercomplex number in the intersection as:

$$i_u i_v = a_{uv,0}+i_1 a_{uv,1}+...+i_{D-1}a_{uv,D-1}, u, v \in \{1, ..., D-1\} \quad (8)$$

The equation (8) determines a multiplication table.

TABLE I
EXAMPLE OF HYPERCOMPLEX MULTIPLICATION TABLE FOR $D = 4$.

$\otimes$	1	i_1	i_2	i_3
1	1	i_1	i_2	i_3
i_1	i_1	-1	i_3	$-i_2$
i_2	i_2	i_3	-1	i_1
i_3	i_3	i_2	$-i_1$	-1

Hypercomplex neural network operators follow the formulations in most cases as (6) and (7). Other operators are obtained in similar ways. For instance, the $\mathbb{C}$ReLU activation is formulated as:

$$\begin{aligned}\mathbb{C}ReLU(\mathbf{x}) = &ReLU(\mathbf{x}_0) + i_1 ReLU(\mathbf{x}_1) + i_2 ReLU(\mathbf{x}_2) \\ &+ ... + i_{D-2}ReLU(\mathbf{x}_{D-2}) + i_{D-1}ReLU(\mathbf{x}_{D-1})\end{aligned} \quad (9)$$

Moreover, hypercomplex operators satisfy the following properties to meet deep interactions in deep neural networks.

Property 1. *For all $\alpha, \beta \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y}, \mathbf{r} \in \mathbb{H}$, we can obtain:*

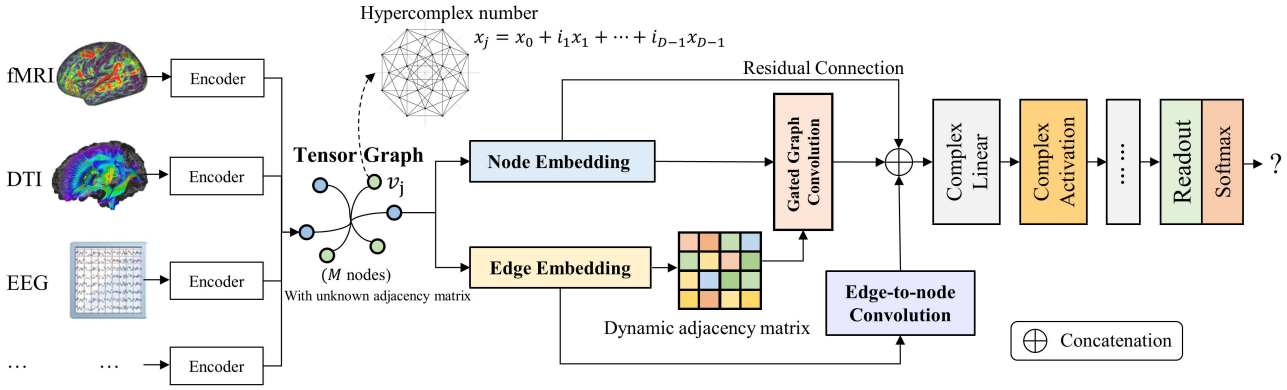

Fig. 2. Illustration of the proposed HC-GNN, with hyercomplex node embedding, hypercomplex edge embedding, and edge-to-node convolution. An encoder is employed to extract deep features from each modality. These modality-specific features are then utilized to construct the hypercomplex nodes.

- $\alpha\mathbf{x} = \alpha\mathbf{x}_0 + i_1(\alpha\mathbf{x}_1) + ... + i_{D-1}(\alpha\mathbf{x}_{D-1}) = \mathbf{x}\alpha$
- $(\alpha\mathbf{x})(\beta\mathbf{y}) = (\alpha\beta)(\mathbf{xy})$
- $\mathbf{x}(\mathbf{y}+\mathbf{r}) = \mathbf{xy} + \mathbf{xr}$ *and* $(\mathbf{x}+\mathbf{y})\mathbf{r} = \mathbf{xr} + \mathbf{yr}$

Property 2. *Given an real-valued loss L, and a complex variable $\mathbf{z} = \mathbf{z}_0 + i_1\mathbf{z}_1 + ... + i_{D-1}\mathbf{z}_{D-1}$, we can obtain:*

$$\nabla_L(z) = \frac{\partial L}{\partial \mathbf{z}} = \frac{\partial L}{\partial \mathbf{z}_0} + i_1\frac{\partial L}{\partial \mathbf{z}_1} + i_2\frac{\partial L}{\partial \mathbf{z}_2} + ... + i_{D-1}\frac{\partial L}{\partial \mathbf{z}_{D-1}} \tag{10}$$

Hypercomplex neural networks offer the ability to model multi-modal inputs by utilizing different imaginary units. For instance, the quaternion domain (with $D = 4$) is suitable for processing 3-modality or 4-modality data, while the octonion domain is suited for more modalities. In practical applications, complex and quaternion neural networks are able to satisfy most multi-modal learning cases.

E. HC-GNN

To aggregate inter-region information and enable graph convolution on tensor graphs, we utilize the hypercomplex convolution operation to embed hypercomplex tensor representations. The structure of the proposed HC-GNN is illustrated in Figure 2. The embedded node representations and adjacency matrix are then used to construct a dynamic graph. Subsequently, we apply gated hypercomplex graph convolution to capture the multi-modal topological information by considering feature interactions within the graph domain.

1) Hypercomplex node embedding: The multi-modal node features are aggregated from the connection features to the centroid node with a hypercomplex multi-layer perception embedding. The output features are obtained by a weighted sum. In detail, given a vertex v_j, the corresponding features x_j are obtained from its connections $\mathcal{N}_j = \{x_{j1}, x_{j2}, ..., x_{jM}\}$:

$$\begin{aligned} h_j = f_n(\mathcal{N}_j) &= \sum_{k\in\mathcal{N}(j)} w_k * x_{jk} \\ &= \sum_{k\in\mathcal{N}(j)} (w_{k,0} + i_1w_{k,1} + i_2w_{k,2} + ... + i_{D-1}w_{k,D-1})* \\ &(x_{jk,0} + i_1x_{jk,1} + i_2x_{jk,2} + ... + i_{D-1}x_{jk,D-1}) \end{aligned} \tag{11}$$

where $w = w_0 + i_1w_1 + i_2w_2 + ... + i_{D-1}w_{D-1}$ is a hypercomplex learnabel parameter for D modalities learning.

2) Dynamic hypercomplex edge embedding: Usually, the edge connections should be conditioned on the node field. While in the brain network study, the derived brain connectivity represents the linear associations to some extent. We use a transformation matrix $\mathbf{W}$ to estimate the dynamic adjacency matrix $\mathbf{L}$ from the connectivity x by:

$$\mathbf{L}_{j,k} = \sigma\sum_{h=1}^{H}\sum_{m=1}^{M} \mathbf{w}_m^{11}\hat{\mathbf{L}}_{j,m} + \mathbf{w}_m^{12}\hat{\mathbf{L}}_{k,m}, \mathbf{L} \in \mathbb{C}^{M\times M} \tag{12}$$

$$\hat{\mathbf{L}}_{h,j,k} = \sum_{m=1}^{M} \mathbf{w}_{m,h}^{21}\hat{\mathbf{x}}_{j,m} + \mathbf{w}_{m,h}^{22}\hat{\mathbf{x}}_{m,k}, \hat{\mathbf{L}} \in \mathbb{C}^{H\times M\times M} \tag{13}$$

where $\mathbf{L}_{j,k}$ represents the j-th row and k-th column element of $\mathbf{L}$, while $\hat{\mathbf{L}}_{h,j,k}$ is the j-th row and k-th column element in the h-th feature map of $\hat{\mathbf{L}}$. $\mathbf{w}^{11}$, $\mathbf{w}^{12}$, $\mathbf{w}^{21}$ and $\mathbf{w}^{22}$ are learned weights, and σ is a sigmoid function.

3) Edge-to-node convolution: The edge-to-node (E2N) convolution is proposed by BrainNetCNN [26], which takes an adjacency matrix from each feature map as an input and outputs a vector. In the hypercomplex operators, the output, $\mathbf{O}$, of an E2N layer is obtained as:

$$\mathbf{O}_{h,j} = \sum_{h=1}^{H}\sum_{m=1}^{M} \mathbf{w}_{m,h}^{31}\hat{\mathbf{L}}_{j,m} + \mathbf{w}_{m,h}^{32}\hat{\mathbf{L}}_{m,j}, \mathbf{O} \in \mathbb{C}^{H\times M} \tag{14}$$

4) Gated Graph Convolution: In this study, we discard the spectral graph convolution, and apply the spatial aggregation to simplify. The hypercomplex graph convolution layer is proposed by replacing the real kernel with hypercomplex-valued kernel. Given a hypercomplex-valued adjacency matrix $\mathbf{L} = \mathbf{L}_0 + i_1\mathbf{L}_1 + ... + i_{D-1}\mathbf{L}_{D-1}$ and an input feature $\mathbf{H} = \mathbf{H}_0 + i_1\mathbf{H}_1 + ... + i_{D-1}\mathbf{H}_{D-1}$, the complex-valued graph convolution is obtained by matrix multiply:

$$\mathbf{S} = GCN(\mathbf{L}, \mathbf{H}; \mathbf{W}) = \mathbf{L} \otimes \mathbf{H}^{l-1} \otimes \mathbf{W} \tag{15}$$

Moreover, the hypercomplex operators inevitably compute and fuse multi-modal representations, which ignore the heterogeneity of inter-modalities. In this regard, we propose to

leverage the gated transformation between different modalities. The gating mechanism has become a common tool in a neural network that facilitates the learning of longer-term relationships and protects the cell state from undesired updates [63]. Intuitively, the gates determine the amount of a signal to pass. In this study, it is implemented by an attention mechanism as an element-wise product and is applied to the graph message passing, resulting in the scaling of the inter-unit aggregation.

$$\begin{aligned} &g_{u,v} = \sigma(\mathbf{W}_{uv,1} \cdot \mathbf{z}_u + \mathbf{W}_{uv,2} \cdot \mathbf{z}_v + \mathbf{b}_{uv}) \\ &\forall u, v \in \{1, 2, ..., D-1\} \end{aligned} \tag{16}$$

where $\mathbf{W}$ and $\mathbf{b}$ are learnable parameters. $g_{u,v}$ denotes gates for the aggregation between the u and v hyperimaginary axes.

And σ is the sigmoid activation functions. Notably, $\mathbf{z}$ represents the transition state representations, where each element $\mathbf{z}(j)$ determines the j-th node by:

$$\mathbf{z}(j) = ||\{Avg_{k \in \mathcal{N}(j)}(x_k), Max_{k \in \mathcal{N}(j)}(x_k)\} \tag{17}$$

$||$ represents the concatenation operation. Accordingly, the gates for inter-axis filtering can be written as a gate matrix $g_\star$ as:

$$\begin{bmatrix} 1 & g_{0,1} & ... & g_{0,D-1} \\ g_{1,0} & 1 & ... & g_{1,D-1} \\ g_{2,0} & g_{2,1} & ... & ... \\ g_{D-1,0} & g_{D-1,1} & ... & 1 \end{bmatrix}$$

To summarize, the hyhercomplex graph convolution layer can be formulated with Eq. (15) and (16) as:

$$\mathbf{H}^l = \sigma[g_\star \circ (\mathbf{L} \otimes \mathbf{H}^{l-1}) \otimes \mathbf{W}] \tag{18}$$

We assume $\mathbf{P} = \mathbf{L} \otimes \mathbf{H}^{l-1} = \mathbf{P}_0 + i_1\mathbf{P}_1 + ... + i_{D-1}\mathbf{P}_{D-1}$, and we can get:

$$g_\star \circ \mathbf{P} = \begin{bmatrix} 1, & ..., & g_{uv} \end{bmatrix} \odot \begin{bmatrix} 1, & ..., & i_u i_v \end{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ ... \\ \mathbf{P}_{D-1} \end{bmatrix} \tag{19}$$

where the $\odot$ represents the element-wise product of the gate matrix $g_\star$ and the hypercomplex multiplication table. For example, when we get $D = 2$ modalities, the gated convolution is formulated as:

$$\begin{aligned} &g_\star \circ \mathbf{L} \otimes \mathbf{H}^{l-1} \\ &= L_0\mathbf{H}_0^{l-1} - L_1\mathbf{H}_1^{l-1} + g_{0,1}i(L_0\mathbf{H}_1^{l-1} + L_1\mathbf{H}_0^{l-1}) \end{aligned} \tag{20}$$

F. Biomarker Interpretation

To inspect the interpretability of our proposed HC-GNN, we employ the saliency maps that have been widely used for biological explanation to identify those important brain regions. In detail, we extract the activation maps of the aggregated graph feature maps $\mathbf{S}$ and performed the statistical CWAS approach, multi-distance multi-variance regression (MDMR) [22] for group-wise comparison.

In detail, a distance matrix in the subject space was calculated for each region. Within each distance matrix, the distance between the saliency maps $\mathbf{S}$ for every possible subject pair among all groups related to region i was calculated by

$$d_{uv}^i = \sqrt{2(1 - r_{uv})} \tag{21}$$

where dis denotes the distance of the saliency maps of subject u and v, and r_{uv} represents the linear correlation between $\mathbf{S}_u$ and $\mathbf{S}_v$ by reference to previous works [22], [24]. A pseudo F-statistic analogous to an F-statistic from a standard ANOVA was performed. The total sum of squares for region i was obtained as

$$SS_T^i = \frac{1}{n}\sum_{u=1}^{n}\sum_{v=u+1}^{n} d_{uv}^{i2} \tag{22}$$

where $n = n_1 + n_2$, the total number of subjects. Meanwhile, the within-group sum of squares was formulated By

$$SS_W^i = \frac{1}{n_1}\sum_{u=1}^{n}\sum_{v=u+1}^{n} d_{uv}^{i2}\varepsilon_{uv}^a + \frac{1}{n_2}\sum_{u=1}^{n}\sum_{v=u+1}^{n} d_{uv}^{i2}\varepsilon_{uv}^b \tag{23}$$

where n_1 and n_2 denote the number of each group respectively. ε_{uv}^a represents the belonging of the subject u and v, which equals to one when u and v within the same group. And then the F-statistic score of the region i would be obtained by

$$F^i = (n-1)\frac{SS_T^i - SS_W^i}{SS_W^i} \tag{24}$$

In addition, a random permutation with 2000 times to subjects was applied to simulate the null distribution, and the pseudo F-statistic score was recomputed for each time. The p-value was finally calculated by counting the pseudo F-statistics from permutated values greater than those derived from the original data. This step was repeated for all ROIs. The Bonferroni correction was applied to control the false positive rate. And the p-value < 0.05 after the correction was determined significant within the experiments.

IV. Experiments

A. Datasets

Our proposed HC-GNN is feasible to model various types of brain networks. In this study, three real-world datasets are employed. All the datasets are enrolled for multi-modal graph classification. The detailed demographic information is listed in Table II with gender (Male/Female), age (Mean±Std), and other phenotypic information (Mean±Std).

1) ADNI Dataset: The ADNI dataset[1] is a longitudinal and multi-modal neuroimaging dataset. We have collected a total of 442 subjects for evaluation, which includes 142 normal controls (NC), 151 individuals with mild cognitive impairment (MCI), and 149 patients diagnosed with Alzheimer's disease (AD). The subjects were selected from various subsites of ADNI, namely ADNI-1, ADN-2/GO, and ADNI-3. Each participant in the study has provided both fMRI and DTI data. To ensure the data integrity and avoid redundancy, we excluded duplicated scans, resulting in a single fMRI and DTI scan per subject being included in the analysis. The subjects were categorized into three groups AD, MCI, and NC based on standard clinical criteria, including assessments such as Mini-Mental State Examination (MMSE) scores and Clinical Dementia Rating (CDR). The primary objective of using the ADNI dataset in this study is to evaluate the coupling between functional and structural connectomes during the progression of Alzheimer's disease.

[1] http://www.adni-info.org/

TABLE II
CHARACTERISTICS OF PARTICIPANTS OF DIFFERENT DATASETS.

Dataset	Task	Modalities	Group	Number	Scale (Mean±Std)	Gender (M/F)	Age (Mean±Std)
ADNI	Task-1	fMRI, DTI	NC	142	-	81/61	75.9±9.7
			MCI	151	-	87/64	77.0±11.7
			AD	149	-	87/62	76.9±8.2
ABIDE-I	Task-2	fMRI (multi-type)	NC	586	-	491/95	16.8±7.7
			ASD	528	-	466/62	17.0$|pm$8.4
ABIDE-II	Task-2 (Testing)	fMRI (multi-type)	NC	677	-	429/248	14.9±9.6
			ASD	559	-	466/93	14.3±8.5
HBN-ADHD	Task-2	fMRI, EEG	NC	105	-	72/33	10.3±3.6
			ADHD	150	-	107/43	9.9±3.0
HBN-Phenotype	Task-3	fMRI, EEG	SWAN	616	0.47±1.04	384/232	10.5±3.1
			SP		4.25±3.81		
			AP		7.91±4.83		

2) ABIDE dataset: The ABIDE database[2] includes T1 structural brain images, resting-state functional MR images, and phenotypic information from 17 different imaging sites. The initial ABIDE-I dataset consists of 528 individuals diagnosed with autism spectrum disorder (ASD) and 586 controls. We employ the ABIDE dataset to evaluate the performance of HC-GNN in studying multiple types of functional connectomes. These connectomes are constructed using various methods, including Pearson coefficient correlation, partial correlation, and tangent space of covariance matrices. To further assess the generalization ability of HC-GNN, we implement the ABIDE-II dataset for testing, which comprises 677 healthy controls and 559 patients with ASD. This dataset allows us to test the effectiveness of HC-GNN on a broader sample population and validate its performance beyond the initial dataset.

3) HBN dataset [64]: The HBN dataset was obtained from a cohort of children and adolescents with age ranging from 5 to 21 years. It includes multiple tasks of functional MRI (fMRI) and electroencephalography (EEG) recordings. For the purpose of this study, we specifically extracted the resting-state fMRI and EEG data to construct the corresponding functional connectomes. Our analysis focused on a total of 616 participants who had both behavioral and cognitive recordings available. The HBN dataset is divided into two distinct parts: HBN-ADHD and HBN-Phenotype. The HBN-ADHD subset consists of 105 healthy control individuals and 150 patients diagnosed with attention-deficit/hyperactivity disorder (ADHD). On the other hand, the HBN-Phenotype cohort comprises 616 participants and includes measurements of various behavioral phenotypes. We utilize this dataset to investigate the coupling between fMRI and EEG functional connectomes and their relationship with behavioral phenotypes and diseases. By examining the functional connectome patterns, we aim to gain insights into the associations between brain connectivity and specific behavioral characteristics or disorders.

Based on the three datasets, our experiments were designed to address three specific tasks:

Task-1: Preclinical Diagnosis of AD: Mild Cognitive Impairment (MCI) is considered a significant stage for the preclinical diagnosis of Alzheimer's disease. To investigate this, we utilized the fMRI and DTI of the ADNI dataset to diagnose the progression of AD.

Task-2: Psychiatric Disease Diagnosis: The ABIDE and HBN datasets provide valuable information on children with autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD), respectively. We extracted and matched patients and healthy controls from these datasets to facilitate the diagnosis of psychiatric diseases by using multi-modal functional connectomes.

Task-3: Behavioral Phenotype Prediction: Psychiatric and learning disorders are prevalent and debilitating conditions throughout the lifespan. Behavioral measures and cognitive tasks play a crucial role in diagnosing these disorders. To predict behavioral phenotypes, we utilized the HBN dataset for the prediction of Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Scale (SWAN, ranging from -3 to 3), Attention Problem (AP, ranging from 0 to 20) scores, and Social Problem (SP, ranging from 0 to 19) scores.

B. Pre-processing

fMRI. All fMRI images underwent pre-processing using the Configurable Pipeline for the Analysis of Connectomes (C-PAC) pipeline [65]. This included several steps such as skull stripping, slice timing correction, motion correction, global mean intensity normalization, nuisance signal regression with 24 motion parameters, and band-pass filtering (0.01-0.08 Hz). The functional images were then registered to the standard anatomical space (MNI152). For region-based analysis, the mean time series for a predefined set of regions of interest (ROIs) were computed and normalized to have zero mean and unit variance. In the ADNI and ABIDE datasets, we used the Schaefer atlas [66] with 100 ROIs to map the fMRI images into parcellations. In the HBN dataset, we utilized the Desikan-Killiany template with 68 ROIs [67] for the cortical cortex learning of both EEG and fMRI data.

DTI. The DTI images were pre-processed by image denoising, head motion, eddy-current, susceptibility distortion, and field inhomogeneity correction by MRtrix-3 [68]. To reconstruct the fiber tracts, we utilized the 2nd order Integration over Fiber Orientation Distributions method [69], resulting in 10

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

million streamlines. Subsequently, a Spherical-deconvolution Informed Filtering of Tractograms technique [70] was applied to reduce the streamline count to 5 million, while preserving the essential structural information. The structural connectivity network was constructed based on the number of streamlines connecting each pair of brain regions. In the ADNI dataset, the DTI images were mapped to parcellations using the Schaefer atlas with 100 ROIs [66].

EEG. The EEG data were recorded in a sound-shielded room at a sampling rate of 500 Hz, using a bandpass filter of 0.1 to 100 Hz. For each participant, the head circumference was measured, and an appropriately sized EEG net was selected. To improve data quality, the EEG signal underwent pre-processing steps. It was filtered using a FIR filter ranging (8-12) Hz which reflects cognitive and memory performance in particular [71]. Noisy channels were identified and removed, and interpolation techniques were applied based on the activity of surrounding channels. To align the EEG data with fMRI space, we performed source reconstruction pre-processing using the Brainstorm software [72]. To analyze the EEG data, the cortical surface was divided into 68 anatomical regions of interest based on the Desikan-Killiany atlas [67]. We focused on the middle 2 minutes of the EEG recording and calculated Pearson correlations by computing correlation coefficients on successive non-overlapping 10-second segments of the EEG time series. The correlation matrices obtained from each segment were averaged to calculate functional connectivity for each participant.

C. Implementation details

In our implementation, we utilize the GRU as the brain connectome encoder that parses the brain network as M sequences with M features. The number of GRU layers is determined through a grid search ranging from 1 to 3. The outputs of the hypercomplex graph convolution layer are then fed into a 3-layer multi-layer perceptron classifier. This classifier is followed by a $\mathbb{C}$ReLU activation function and a dropout layer to enhance model generalization. To train the models, we set the learning rate as 3e-4 and the weight decay as 5e-5. All models in this study are trained for 400 epochs, and early stopping is employed if the loss does not decrease for 50 epochs. The training process is performed using PyTorch on a single NVIDIA 2080-Ti GPU. For evaluation, we employ a 10-fold cross-validation strategy, where 10% of the samples are randomly selected for testing in each fold. In the disease diagnosis experiments (Task-1 and Task-2), we assess the performance based on diagnosis accuracy (Acc), sensitivity (Sen), and specificity (Spe). In the behavioral phenotype prediction experiments (Task-3), the performance is evaluated using mean absolute error (MAE), mean squared error (MSE), and Pearson correlation coefficient (PCC).

D. Competitive methods

In this study, we compare our proposed HC-GNN with baseline machine learning approaches and well-estimated graph methods. These methods include:

Machine learning baselines. In order to establish a baseline for classification, we compare our proposed method against conventional machine learning approaches, the support vector machine (**SVM**) and multi-layer perceptron (**MLP**). To obtain a classification score for each subject, we input the upper matrix of the brain networks into these classifiers. For the MLP, we explore different layer settings ranging from 1 to 4.

Tensor-based approaches. For the multi-modal data learning, tensor-based approaches such as CP and Tucker methods would extract key features for classification. In our evaluation, we employed these tensor-based methods, referred to as **TD-MLP** and **CP-MLP**, respectively. These methods involved matrix decomposition techniques to extract the principal features from the tensors. The extracted features were then vectorized and fed into a multi-layer perceptron (MLP) for classification. Similarly, the TD-MLP method utilized tensor decomposition to obtain essential features, which were subsequently fed into an MLP for classification purposes.

Convolutional Nerual Networks. BrainNetCNN [26] was first proposed to tackle the brain graph networks as grid-like data, which achieves state-of-the-art performances in brain network studies. BrainNetCNN was implemented by multiple convolution layers, where various types of brain connectivity matrices are concatenated by channel.

Supervised Graph Neural Networks. BrainGNN [28] introduces ROI-aware graph convolutional layers and ROI-selection pooling layers for neurological biomarker prediction. In our implementations, the original partial Pearson and full correlation coefficients were replaced by multi-modal brain networks, and the hyperparameters are searched in a grid search. Moreover, **M-GCN** [39] focuses on aggregating functional representations by incorporating regularization with the structural graph Laplacian. The **HGNN** (Hypergraph Graph Neural Network) [43] leverages hyperstructure to encode multi-modal connectivity through hyperedges. The **DHGNN** (Dynamic Graph Hyper-Neural Network) [44] extends the HGNN framework to handle dynamic graphs. For these three models, we implemented them based on the original proposed architecture as described in their respective papers.

Semi-supervised Graph Neural Networks. Transductive Graph Neural Networks (GNNs) offer valuable capabilities for studying medical neuroimages by semi-supervision, especially in scenarios with limited sample sizes. In **population-GCN**, a population graph is built, where each node is represented by concatenating the vectorized upper matrix of the brain networks of a subject. Key features were selected by recursive feature elimination and vectorized into a set for each vertex and then concatenated. The number of selected features is searched. The number of selected features is determined through a search process with a step of 200. The adjacency matrix is constructed using both phenotype values including gender and age and the similarity between node features. In **MMP-GCN** [46], 200 multi-modal features are selected for the classification of the transductive graphs. In addition, we discarded the multi-center information for the ADNI and HBN datasets, and applied the gender and age information to construct the graph.

Others. The Triplet Attention Network (**TAN**) [47] is a

TABLE III
CLASSIFICATION PERFORMANCE ON THE **TASK-1: PRECLINICAL DIAGNOSIS OF AD** IN TERMS OF ACCURACY (ACC %), SENSITIVITY (SEN %), AND SPECIFICITY (SPE %) AMONG ALL COMPETITIVE MODELS.

Model	Type	NC vs MCI			MCI vs AD			NC vs AD		
		Acc ↑	Sen ↑	Spe ↑	Acc ↑	Sen ↑	Spe ↑	Acc ↑	Sen ↑	Spe ↑
SVM	ML	52.2±9.3	52.4±9.0	52.2±10.6	60.0±9.9	59.8±14.9	60.8±7.8	75.4±9.8	77.5±14.0	75.3±8.8
MLP	ML	70.1±8.8	75.0±15.8	73.1±13.6	69.1±5.4	73.6±9.3	70.1±9.9	77.8±6.5	76.6±6.8	80.4±8.3
CP-MLP	Tensor, ML	73.9±10.2	77.2±10.1	75.8±16.1	75.6±7.2	77.4±10.8	75.9±7.8	75.2±8.1	78.7±9.5	73.6±7.9
TD-MLP	Tensor, ML	73.85±7.6	73.9±9.5	76.6±9.5	73.2±5.9	79.8±12.7	70.7±5.4	75.9±10.9	82.6±14.7	73.2±10.6
BrainNetCNN [26]	CNN	72.08±7.1	75.9±9.3	70.4±10.8	71.8±5.0	74.4±6.8	70.8±5.0	81.3±11.3	81.0±11.7	82.1±10.1
BrainGNN [28]	GNN	76.0±5.9	81.1±9.9	71.5±12.6	74.3±4.8	76.6±5.2	73.5±12.8	82.0±2.4	84.6±5.6	76.8±6.1
M-GCN [39]	GNN	75.7±6.3	80.1±4.8	72.9±6.1	75.8±6.9	76.1±8.1	75.0±9.0	84.4±5.7	84.1±6.0	85.6±7.1
HGNN [43]	GNN	78.1±6.6	80.2±8.4	74.7±9.4	76.8±5.5	83.5±12.8	76.0±7.5	81.4±8.0	82.5±6.1	82.0±12.1
DHGNN [44]	GNN	79.2±6.0	83.0±9.9	77.6±7.3	75.8±4.2	85.4±12.9	73.9±8.4	82.5±4.1	86.4±10.7	82.1±11.4
Population-GCN [32]	semi-GNN	75.5±5.9	82.5±10.0	71.6±8.0	75.5±5.9	80.2±14.4	74.0±5.4	82.0±7.5	85.9±7.5	75.9±8.0
MMP-GCN [46]	semi-GNN	80.6±7.6	86.2±10.2	79.5±10.1	79.6±11.0	82.52±14.0	**82.2±14.3**	84.4±11.1	88.1±12.4	85.1±13.2
TAN [47]	Attention	80.2±5.4	82.0±11.1	79.2±11.6	81.2±3.5	83.7±8.8	79.0±9.7	85.9±4.5	89.3±12.9	86.3±10.3
HC-GNN (Ours)	Tensor, GNN	**81.5±4.1**	**90.6±10.0**	**80.7±13.4**	**82.5±6.3**	**88.0±8.7**	**82.2±12.6**	**87.5±8.4**	**91.9±6.5**	**88.0±11.3**

cutting-edge network specifically designed for brain disorder diagnosis. TAN adopts a Transformer-like architecture that incorporates self-attention and MLP for effective representation learning. In our study, the hidden size is set as 32, and 2 self-attention encoders with 4 heads are implemented. The dropout rate is set as 0.4.

V. RESULTS

A. Evaluation results on different tasks and datasets

1) Classification on the progression of AD: The quantitative results for the three classification sub-tasks of Task-1 (NC vs. MCI, MCI vs. AD, and NC vs. AD) are presented in Table III, with the average and standard deviation across folds. Overall, tensor-based approaches (CP-MLP and TD-MLP) demonstrate promising performances across all three classification sub-tasks compared with MLP and SVM. These methods provide simple and efficient ways to model multi-modal data. In comparison to machine learning methods, deep learning models, especially graph neural networks, consistently enhance classification performance, such as M-GCN and HGNN. This suggests that the structural topological properties of the brain are valuable for distinguishing the preclinical AD. Besides, the semi-supervised GNN method, MMP-GCN, outperforms most supervised GNN methods. However the condition is opposite in Population-GCN. This indicates that such transductive graph neural networks are still limited by the way of constructing the graph adjacency matrix and node features, which limits their generalizability to different tasks. Furthermore, when considering the three classification sub-tasks, our proposed HC-GNN emerges as the top-performing approach among all models, achieving accuracies of 81.5%, 82.5%, and 87.5% in distinguishing NC from MCI, MCI from AD, and NC from AD, respectively. This outstanding performance can be attributed to the deep intersection among modalities of the hypercomplex graph network.

2) Classification of the psychiatric disease diganosis: Table IV presents the comparison results for distinguishing ASD and ADHD from NC on the ABIDE-I and HBN datasets, respectively. The table includes the mean and standard deviation across folds, with the best results highlighted in bold and the second-best results underlined. Upon analyzing the results, we observe that, similar to the ADNI dataset, GNN approaches generally perform the best in most cases. Interestingly, tensor-based methods achieve promising results and outperform some GNN models. For instance, on the ABIDE-I dataset, CP-MLP and TD-MLP exhibit better accuracy performances (CP-MLP: 71.1%, TD-MLP: 71.3%) compared to BrainGNN (69.7%), M-GCN (68.7%), and MMP-GCN (70.5%). On one hand, BrainGNN and M-GCN were available to handle two modalities, and their performance suffers when dealing with limited connectome information. On the other hand, although MMP-GCN befinets from semi-supervised learning on a transductive graph, it fails to capture inter-modal dependencies effectively. In contrast, our proposed HC-GNN and TAN are capable of modeling intersection representations between modalities, thereby exhibiting better performance. Furthermore, on the HBN-ADHD dataset, our proposed HC-GNN outperforms the second-best model, TAN, with a 3% improvement in accuracy. This is due to the fact that most compared models are designed for either fMRI connectomes or DTI connectomes and struggle to generalize well to the learning of both fMRI and EEG. In contrast, our HC-GNN exhibits better scalability when learning from various types of brain networks. We suspect that this is caused by the deep intersection among modalities that facilitates generating high-order generalizations of matrices for multi-modal learning.

3) Behavioral phenotype prediction: Furthermore, we conducted behavioral phenotype prediction for attention problem score, social problem score, and the Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Scale score. The quantitative results for these sub-tasks are presented in Table V, including metrics such as MAE, MSE, and Pearson coefficient correlation. Overall, we observed low linear correlation (PCC) ranging from 0.1 to 0.3 between the three phenotypes and the predicted scores. This can be attributed to the fact that the fMRI and EEG recordings used in this study are based on resting-state paradigms, which may not strongly correlate with specific behavioral phenotypes. Nonetheless, our HC-GNN consistently improves the linear correlation, achieving values of around 0.3 across all three sub-tasks. Furthermore,

TABLE IV
CLASSIFICATION PERFORMANCE ON THE **TASK-2: PSYCHIATRIC DISEASE DIAGNOSIS** IN TERMS OF ACCURACY (ACC %), SENSITIVITY (SEN %), AND SPECIFICITY (SPE %) AMONG ALL COMPETITIVE MODELS. * INDICATES THAT THE MODELS ARE TRAINED WITH AND ONLY AVAILABLE FOR TWO MODALITIES.

Model	Type	NC vs ASD (ABIDE-I)			NC vs ADHD (HBN-ADHD)		
		Acc ↑	Sen ↑	Spe ↑	ACC ↑	Sen ↑	Spe ↑
SVM	ML	69.0±3.7	71.3±3.3	67.8±4.3	60.0±4.3	60.1±2.9	63.9±16.2
MLP	ML	70.1±2.2	68.8±2.1	71.7±3.8	69.5±2.5	69.9±4.7	72.9±9.3
CP-MLP	Tensor, ML	71.1±3.1	69.4±4.7	72.3±4.4	72.9±5.4	73.9±5.7	74.2±11.8
TD-MLP	Tensor, ML	71.3±2.2	74.0±7.5	70.0±2.1	73.8±5.4	72.2±6.6	78.2±2.6
BrainNetCNN [26]	CNN	72.6±2.5	72.5±3.8	**73.6±4.5**	69.9±7.6	67.9±7.7	**81.0±7.1**
BrainGNN* [28]	GNN	69.7±2.8	66.7±9.9	72.5±8.3	75.2±4.0	**88.6±8.5**	73.9±7.2
M-GCN* [39]	GNN	68.7±6.8	68.7±6.0	72.9±4.9	74.2±4.0	80.0±2.6	74.7±6.0
HGNN [43]	GNN	71.4±2.6	70.2±4.2	73.0±3.9	73.8±5.8	84.6±10.2	72.1±5.1
DHGNN [44]	GNN	71.9±3.3	72.1±3.5	72.0±4.4	72.5±4.0	85.6±8.0	70.2±4.2
Population-GCN [32]	semi-GNN	69.9±1.4	**79.0±5.4**	68.0±1.8	76.2±10.1	79.3±7.1	73.1±9.8
MMP-GCN [46]	semi-GNN	70.5±3.9	71.1±9.6	72.4±7.9	76.7±5.1	76.1±6.3	75.6±3.7
TAN* [47]	Attention	71.7±2.8	71.6±2.7	69.5±5.3	78.6±5.0	83.7±8.1	75.7±5.8
HC-GNN (Ours)	Tensor, GNN	**73.6±3.6**	76.2±4.4	73.1±7.2	**81.6±6.2**	87.3±4.8	78.7±8.2

TABLE V
CLASSIFICATION PERFORMANCE ON THE **TASK-3: BEHAVIRAL PHENOTYPE PREDICTION** IN TERMS OF ACCURACY (ACC %), SENSITIVITY (SEN %), AND SPECIFICITY (SPE %) AMONG ALL COMPETITIVE MODELS. * INDICATES THAT THE MODELS ARE ONLY AVAILABLE FOR TWO-MODALITY DATA LEARNING AND TRAINED WITH TWO MODALITIES.

Model	Type	HBN-SWAN			HBN-SP			HBN-AP		
		MAE ↓	MSE ↓	PCC ↑	MAE ↓	MSE ↓	PCC ↑	MAE ↓	MSE ↓	PCC ↑
SVM	ML	0.804±0.083	1.103±0.237	0.119±0.105	2.966±0.339	14.903±3.702	0.115±0.121	4.035±0.444	23.224±4.340	0.121±0.107
MLP	ML	0.748±0.088	0.996±0.208	0.209±0.066	2.872±0.320	14.833±3.588	0.206±0.100	3.907±0.445	23.119±4.253	0.196±0.064
CP-MLP	Tensor, ML	0.780±0.093	1.062±0.225	0.146±0.076	2.883±0.349	14.706±4.211	0.269±0.135	3.810±0.381	22.560±3.201	**0.286±0.083**
TD-MLP	Tensor, ML	0.762±0.083	1.033±0.210	0.206±0.104	2.878±0.329	14.688±3.603	0.222±0.119	3.857±0.445	23.407±4.426	0.233±0.088
BrainNetCNN [26]	CNN	0.766±0.085	1.052±0.216	0.192±0.121	2.798±0.333	14.621±3.609	0.237±0.113	3.826±0.457	22.249±4.334	0.261±0.069
BrainGNN [28]	GNN	0.791±0.081	1.074±0.210	0.151±0.083	3.028±0.458	16.322±4.620	0.141±0.096	4.054±0.452	23.300±4.489	0.177±0.103
M-GCN [39]	GNN	0.771±0.085	1.050±0.213	0.228±0.124	2.883±0.340	14.575±3.646	0.197±0.099	**3.807±0.414**	**22.011±4.167**	**0.286±0.134**
HGNN [43]	GNN	0.746±0.082	1.003±0.217	0.295±0.092	**2.781±0.360**	**14.512±4.072**	**0.290±0.112**	3.836±0.393	22.537±3.362	0.213±0.111
DHGNN [44]	GNN	0.750±0.078	1.027±0.214	0.241±0.059	3.077±0.318	18.291±3.677	0.153±0.100	3.945±0.720	28.008±10.799	0.186±0.100
Population-GCN [32]	semi-GNN	0.797±0.090	1.099±0.236	0.205±0.333	2.975±0.366	15.863±4.290	0.126±0.318	4.262±0.712	31.326±21.460	0.197±0.288
MMP-GCN [46]	semi-GNN	0.617±0.082	0.950±0.680	0.199±0.130	2.867±0.392	14.799±4.477	0.163±0.224	3.998±0.380	22.790±4.741	0.205±0.199
TAN [47]	Attention	0.602±0.148	0.663±0.678	0.188±0.179	3.001±0.424	16.216±2.060	0.216±0.205	3.868±0.356	22.854±4.343	0.199±0.146
HC-GNN (Ours)	Tensor, GNN	**0.486±0.039**	**0.642±0.099**	**0.322±0.121**	2.795±0.325	14.782±3.315	0.281±0.119	3.809±0.428	22.607±4.974	0.282±0.075

our HC-GNN outperforms other methods in predicting SWAN scores significantly. However, it performs slightly worse than HGNN in measuring SP and M-GCN in measuring AP. This is a limitation of our proposed HC-GNN, as it is not able to consistently achieve the outstanding performances on regression tasks. This limitation may arise due to the deep intersection of multiple modalities. When different modalities exhibit significant variations in prediction, the overall performance can be hindered by the underperforming modality, even with the inclusion of a gating mechanism.

B. Ablation study

In this study, we focus on utilizing the gating mechanism in our HC-GNN based on a hypercomplex neural network to effectively model multi-modal tensor graphs. In this section, we present our ablation studies that aim to investigate the effectiveness of the gating mechanism, the modeling of HNN, and the selection of hypercomplex activation functions in the learning of brain connectomes. These studies provide valuable insights into the contribution of each component and their impact on the overall performance of our HC-GNN model.

1) Gating Mechanism: To examine the effectiveness of the gating mechanism, we conducted a comparison between our proposed HC-GNN models with and without the gating mechanism (referred to as "w Gating" and "w/o Gating" respectively). The results, depicted in Figure 3, showcase the performance in terms of accuracy for Task-1 and Task-2, as well as the mean absolute error for Task-3. It is evident that across all three tasks, our HC-GNN with the gating consistently outperforms the model without the gating, which indicates that the deep intersection of hypercomplex neural networks introduces unwanted inductive noise in coupling multi-modal brain networks. However, it is important to note that the gating does not always work in every case. Specifically, in Task-3, the gating does not significantly enhance the prediction of SWAN and SP. This could be attributed to the fact that the gating mechanism, learned through the attention paradigm, may not always facilitate sparsely-distributed representations, leading to an averaged and potentially uninformative attention mapping in certain scenarios. Additionally, the intersection between heterogeneous modalities may have minimal impact on the behavioral phenotype prediction. Nonetheless, it is worth highlighting that even in cases where the gating mech-

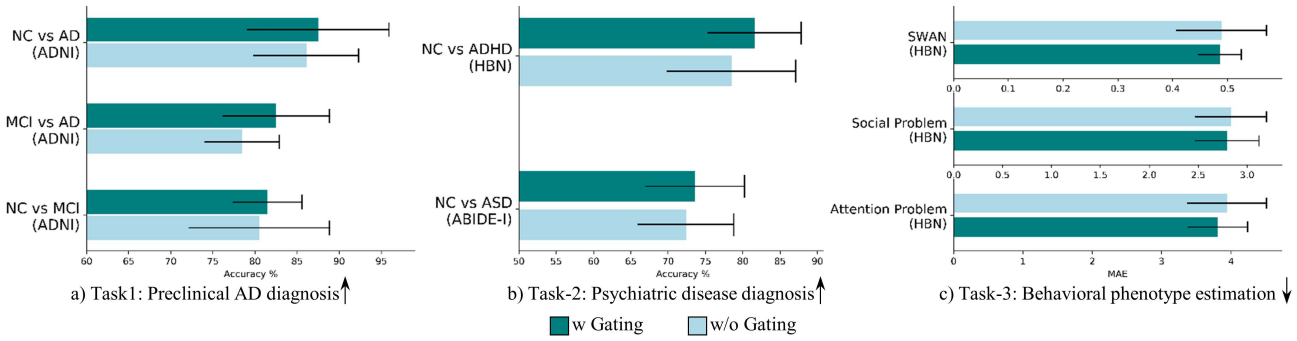

Fig. 3. Ablation studies on the gating mechanism on all three tasks. Accuracy and MAE are used as the measurements for evaluating the classification (Task-1 and Task-2) and theregression tasks (Task-3). The green and blue bar plots indicate HC-GNN with and without the gating respectively.

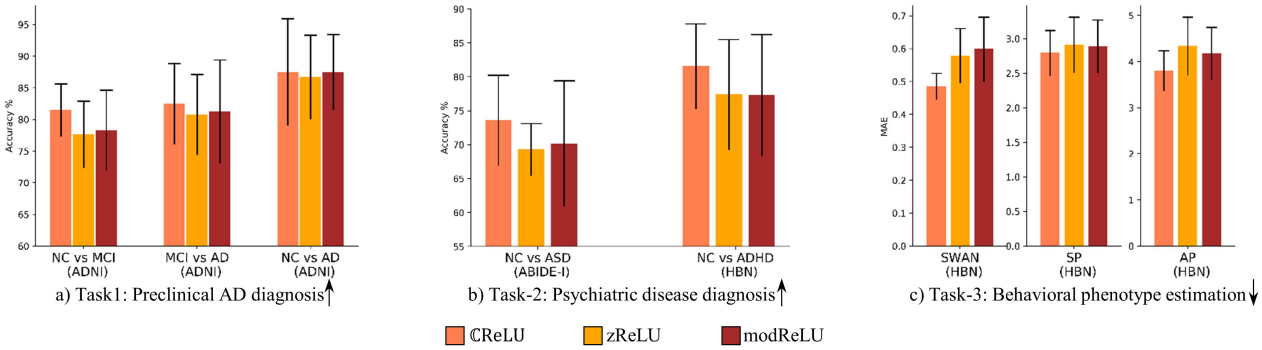

Fig. 4. Ablation studies on the hypercomplex activation function on all three tasks. Accuracy and MAE are used as the measurements for evaluating the classification (Task-1 and Task-2) and theregression tasks (Task-3). The orange, yellow and brown bar plots indicate HC-GNN is trained by ℂReLU, zReLU, and modReLU respectively.

anism is not particularly influential, it does not diminish the overall performance. Based on these observations, we consider the gating mechanism to be a powerful tool for facilitating intersection between modalities.

2) Activation Function: Furthermore, we conducted an evaluation of the hypercomplex activation functions, including ℂReLU, zReLU, and modReLU. The comparative results presented in Table 4 demonstrate that the ℂReLU function consistently achieves the best performance across all tasks. This is reasonable for the hypercomplex neural networks, as it ensures compliance with the Cauchy-Riemann equations. We hypothesize that modReLU and zReLU treat the hypercomplex inputs as a whole and filter the feature maps accordingly. In our study, however, the multi-modal inputs still retain their separate information within distinct imaginary units to some extent. Consequently, the ℂReLU function emerges as a more suitable choice for effectively mapping the heterogeneous modalities, leveraging the real and imaginary parts separately.

3) Impact of Hypercomplex neural network: In this study, our approach involves constructing tensor-graphs from multi-modal data and employing hypercomplex neural networks to analyze these tensor-graphs. A crucial aspect is to validate the effectiveness of using hypercomplex neural networks to model tensor-graphs. To this end, we conducted ablation studies comparing two approaches: (1) Directly using multi-channel real-valued graph neural networks, and (2) extracting key features through Tucker Decomposition and feeding them into a multi-channel real-valued neural network. The multi-channel features are flattened before performing message passing on graphs. Table VI presents the comparative results across all three tasks, with the accuracy for Task-1 and Task-2, and the MAE for Task-3. It can be observed that the performances of the real-valued graph neural networks are on par with those of BrainNetCNN in Table III, Table IV and Table V, as both models capture inter-modal dependencies using multi-channels. However, the use of Tucker Decomposition consistently improves performance, as it enables the extraction of key components and treats the multi-modal matrix as a unified entity. Nonetheless, our proposed HC-GNN consistently outperforms these methods. Notably, in predicting SWAN, the real-valued neural networks achieve comparable performance with other GNN models such as BrainNetCNN, BrainGNN, and M-GCN. In contrast, our proposed hypercomplex paradigm further reduces the MAE error from 0.760 to 0.486, showcasing its superior performance.

C. Model generalization ability

To assess the model's generalization ability due to limited samples in neuroimaging and EEG datasets, we conducted testing using the ABIDE-II dataset. The results of this evaluation

TABLE VI
ABLATION STUDIES ON THE HYPERCOMPLEX MODELING OF TENSOR-GRAPHS AMONG ALL THREE TASKS IN TERMS OF ACCURACY (ACC %) AND MEAN ABSOLUTE ERROR (MAE).

Dataset	ADNI			ABIDE-I	HBN			
Task ID	Task-1			Task-2		Task-3		
Task Name	NC vs. MCI	MCI vs. AD	NC vs. AD	NC vs. ASD	NC vs. ADHD	SWAN	SP	AP
Metric	ACC ↑	ACC ↑	ACC ↑	ACC ↑	ACC ↑	MAE ↓	MAE ↓	MAE ↓
Real-valued	77.5±6.6	74.5±10.5	82.2±10.2	70.5±7.8	74.3±11.9	0.792±0.076	2.885±0.337	3.921±0.488
Real-valued,TD	80.0±8.7	80.6±9.3	85.8±13.3	71.0±2.6	76.6±5.5	0.780±0.081	2.901±0.363	3.951±0.420
Hypercomplex (Ours)	**81.5±4.1**	**82.5±6.3**	**87.5±8.4**	**73.6±6.6**	**81.6±6.2**	**0.486±0.039**	**2.795±0.325**	**3.809±0.428**

are presented in Table VII, which includes the average and deviation results obtained from the 10-fold models trained on the ABIDE-I dataset. Comparing the results of the ABIDE-I dataset from Table IV, we can observe that the second-best model in the ABIDE-I dataset, BrainNetCNN, performs poorly in the ABIDE-II dataset. This indicates that BrainNetCNN has limited generalization ability for predicting ASD with different datasets. On the other hand, TAN, which utilizes self-attention to model multi-modal representations, demonstrates better generalization ability. This finding shows evidence that Transformer-like architecture is able to generalize to various fields. Compared with these models, our proposed HC-GNN achieves the best generalizability among all the models, outperforming the second best model with 1.6% improvements.

D. Model Interpretability

To better understand which brain regions contribute most to the classification of the brain disorders, we analyzed the saliency maps to identify the brain regions that contribute most significantly. Figure 5 displays the F-statistic values of the regions that were found to be statistically significant (with a p-value of less than 0.05) during the group-wise comparison.

In the comparison between individuals with NC and those with MCI, three key regions were identified with significant differences. These regions include the left temporal lobe, the left parietal cortex, and the right prefrontal cortex (PFC) of the Default Mode Network (DMN). Besides the temporal lobe, in the distinction between AD and MCI, the right precentral ventral cortex of the Dorsal Attention Network (DAN) was also found to be discriminative. Additionally, in the comparison between NC and AD, significant differences were observed in the cingulate regions of both sides of the Frontoparietal Network (FN), the left temporal lobe of the DMN, and the left prefrontal cortex of the DMN. The temporal lobes play a crucial role in declarative memory and recognition. The posterior cingulate cortex (PCC) and medial temporal cortex are anatomically and functionally connected, and amyloid deposition in these regions contributes to synaptic dysfunction and eventual neuronal loss in the hippocampus and surrounding cortical structures, leading to memory impairment [73], [74]. Furthermore, AD is a neurodegenerative condition that affects heteromodal association regions, such as PFC. Dysfunction in PFC, which is involved in semantic memory processing, has been observed in AD patients, suggesting compromised semantic memory in this population [75], [76]. Our findings align with these studies and provide additional evidence supporting their conclusions.

TABLE VII
MODEL GENERALIZABILITY OF CLASSIFICATION PERFORMANCE ON INDEPENDENT TESTING DATASET, ABIDE-II, IN TERMS OF ACCURACY (ACC %), SENSITIVITY (SEN %), AND SPECIFICITY (SPE %) AMONG ALL COMPETITIVE MODELS. * INDICATES THAT THE MODELS ARE TRAINED WITH AND ONLY AVAILABLE FOR TWO MODALITIES.

Model	Type	NC vs ASD (ABIDE-II)		
		Acc ↑	Sen ↑	Spe ↑
SVM	ML	66.1±0.4	63.7±0.7	67.8±0.3
MLP	ML	67.3±0.3	66.3±1.5	68.1±1.4
CP-MLP	Tensor, ML	64.1±0.5	59.1±0.9	67.2±0.9
TD-MLP	Tensor, ML	65.1±0.4	65.4±0.8	65.3±0.4
BrainNetCNN [26]	CNN	66.2±0.7	66.0±1.6	66.5±1.5
BrainGNN* [28]	GNN	63.2±0.6	50.1±3.4	74.1±5.6
M-GCN* [39]	GNN	64.8±0.1	64.8±2.2	64.8±0.1
HGNN [43]	GNN	65.4±0.7	63.9±1.9	66.5±1.4
DHGNN [44]	GNN	64.6±0.7	61.8±0.8	66.3±1.1
Population-GCN [32]	semi-GNN	66.0±0.3	66.2±4.0	66.3±0.4
MMP-GCN [46]	semi-GNN	66.3±0.2	**68.6±0.4**	59.6±0.2
TAN* [47]	Transformer	67.1±0.2	66.3± 0.7	68.4±0.1
HC-GNN (Ours)	Tensor, GNN	**68.7±0.5**	66.9±2.7	**68.6±1.7**

Furthermore, significant differences were observed in PFC of the Default Mode Network and PCC of the Frontoparietal Network in the comparison between individuals with NC and ASD. And the right middle temporal cortex, precuneus, and left superior temporal cortex are key regions in the comparison between NC and ADHD. These regions predominantly belong to the default mode network, which is involved in adaptive responses in both ASD and ADHD [77]–[79]. Our findings in the prefrontal and temporal cortex coincides with these studies. Besides, previous studies suggest that altered precuneus connectivity is involved in the neuropathology of ADHD [80], [81], which is consistent with our results.

VI. DISCUSSION

In this study, we observed that the classification results on the ADNI dataset exhibit relatively large deviations, whereas the deviations are smaller in the ABIDE-I dataset. This difference can be attributed to the larger size of the cohort in the ABIDE-I dataset. It's worth noting that the ADNI dataset comprises participants with both fMRI and DTI images, which are challenging to collect. Although the integration of functional and structural connectomes has greatly facilitated computer-aided diagnosis of neurodegenerative diseases, it comes with a high cost in terms of data acquisition. One approach to

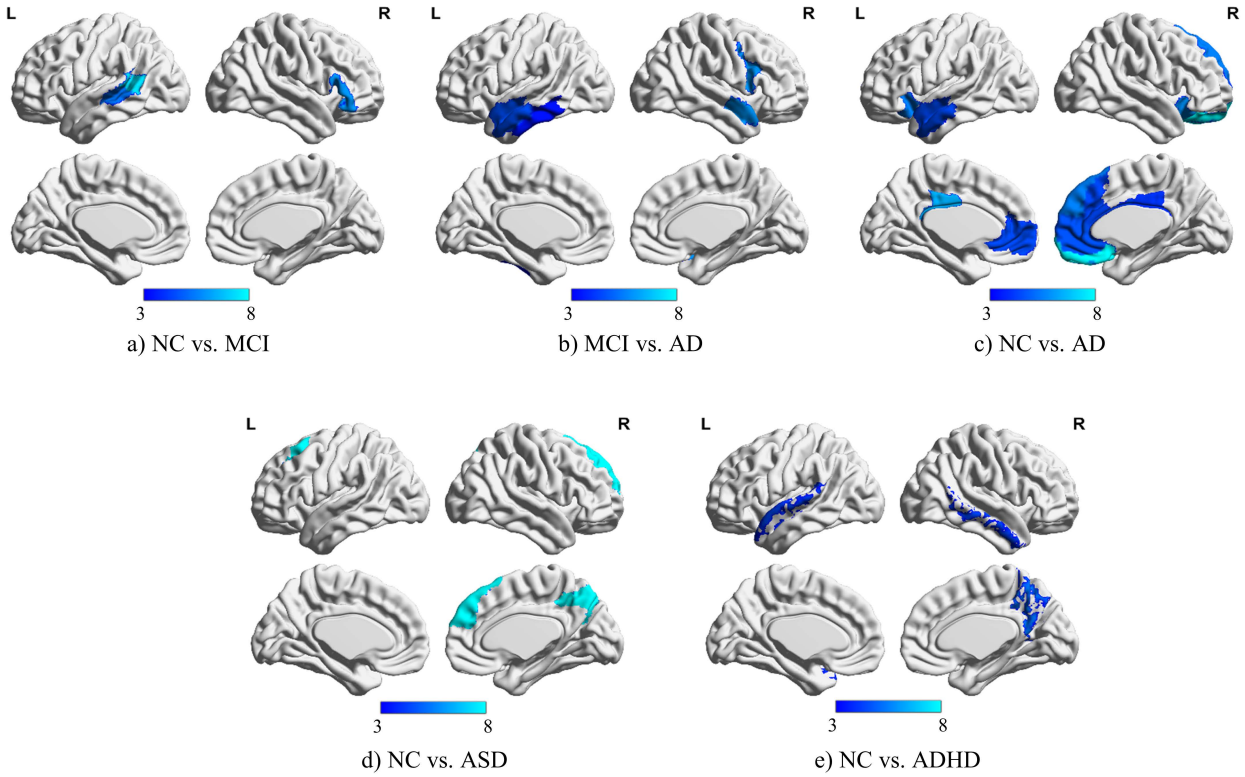

Fig. 5. Group-wise comparison on (a) NC vs. MCI, (b) MCI vs. AD, (c) NC vs. AD, (d) NC vs. ASD, and (e) NC vs. ADHD. Statistical results (F-statistic values are shown in blue with the cutoff ranging from 3 to 8) on the saliency maps with only significant regions (adjusted p-value < 0.05) are displayed.

mitigate this challenge is to utilize other types of modalities. The results from HBN-ADHD demonstrate that incorporating resting-state fMRI and EEG data is available for diagnosing ADHD with high accuracy. EEG data is more easily accessible compared to MRI scans. Furthermore, our HC-GNN outperforms previous studies [38], [82] (around 71% accuracy) in terms of classification results for distinguishing between NC and ADHD. While these previous studies primarily rely on fMRI, which limits their performance, our HC-GNN achieves an accuracy of 81.6%. In conclusion, our HC-GNN provides novel insights in modelling various types of brain networks in a promising performance.

In addition to its strengths, our proposed HC-GNN also has certain limitations. On one hand, the deep intersection operations involved in hypercomplex neural networks are time-consuming, even though they do reduce the number of parameters compared to standard convolutional layers in real-world problems. While this issue is acceptable for brain network studies given their relatively small sample sizes, it remains a time-consuming process for other multi-modal cases such as images and language texts. On the other hand, as discussed in the results, the performance of HC-GNN heavily relies on the consistency of performances among various modalities. If there are significant variations in prediction performances across different modalities, the learning performance of the multi-modal model can be hindered by the underperforming modality. In our future work, we plan to explore strategies to optimize the training process and optimize the network in other multi-modal applications.

VII. Conclusion

In this study, we propose to implement hypercomplex neural network for modeling the multi-modal tensor-graphs. And we leverage the gating and dynamic graph mechanisms for the dynamic coupling of multi-modal dependencies. Through extensive experiments, we demonstrate the significant potential of our HC-GNN in modeling multi-modal brain networks, showcasing strong generalizability and scalability. We evaluated our approach using various types of brain networks, highlighting its superior performance compared to state-of-the-art methods in brain disease diagnosis and promising phenotype estimation results. Our HC-GNN provides a novel way of coupling multiple types of brain networks and facilitates the deep intersection of multi-modal connectome for representation learning.

References

[1] T. Y. Lee, W. H. Jung, Y. B. Kwak, Y. B. Yoon, J. Lee, M. Kim, E. Kim, and J. S. Kwon, "Distinct neural networks associated with obsession and delusion: a connectome-wide association study," *Psychological medicine*, vol. 51, no. 8, pp. 1320–1328, 2021.

[2] A. Sharma, D. H. Wolf, R. Ciric, J. W. Kable, T. M. Moore, S. N. Vandekar, N. Katchmar, A. Daldal, K. Ruparel, C. Davatzikos *et al.*, "Common dimensional reward deficits across mood and psychotic disorders: a connectome-wide association study," *American Journal of Psychiatry*, vol. 174, no. 7, pp. 657–666, 2017.

[3] J. Pu, J. Wang, W. Yu, Z. Shen, Q. Lv, K. Zeljic, C. Zhang, B. Sun, G. Liu, and Z. Wang, "Discriminative structured feature engineering for macroscale brain connectomes," *IEEE Transactions on Medical Imaging*, vol. 34, no. 11, pp. 2333–2342, 2015.

[4] J. D. Gabrieli, S. S. Ghosh, and S. Whitfield-Gabrieli, "Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience," *Neuron*, vol. 85, no. 1, pp. 11–26, 2015.

[5] C. Li, M. Liu, J. Xia, L. Mei, Q. Yang, F. Shi, H. Zhang, and D. Shen, "Predicting brain amyloid-β pet grades with graph convolutional networks based on functional mri and multi-level functional connectivity," *Journal of Alzheimer's Disease*, no. Preprint, pp. 1–15, 2022.

[6] S. Graham, J. Jiang, V. Manning, A. B. Nejad, K. Zhisheng, S. R. Salleh, X. Golay, Y. I. Berne, and P. J. McKenna, "Iq-related fmri differences during cognitive set shifting," *Cerebral Cortex*, vol. 20, no. 3, pp. 641–649, 2010.

[7] Y. Yang, C. Ye, and T. Ma, "A deep connectome learning network using graph convolution for connectome-disease association study," *Neural Networks*, vol. 164, pp. 91–104, 2023.

[8] Y. Yang, X. Guo, Z. Chang, C. Ye, Y. Xiang, and T. Ma, "Multi-modal dynamic graph network: Coupling structural and functional connectome for disease diagnosis and classification," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 1343–1349.

[9] 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.

[10] E. T. Bullmore and D. S. Bassett, "Brain graphs: graphical models of the human brain connectome," *Annual review of clinical psychology*, vol. 7, pp. 113–140, 2011.

[11] A. Zalesky and A. Fornito, "A dti-derived measure of cortico-cortical connectivity," *IEEE transactions on medical imaging*, vol. 28, no. 7, pp. 1023–1036, 2009.

[12] S. I. Cunningham, D. Tomasi, and N. D. Volkow, "Structural and functional connectivity of the precuneus and thalamus to the default mode network," *Human Brain Mapping*, vol. 38, no. 2, pp. 938–956, 2017.

[13] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J.-P. Thiran, R. Meuli, and P. Hagmann, "Predicting human resting-state functional connectivity from structural connectivity," *Proceedings of the National Academy of Sciences*, vol. 106, no. 6, pp. 2035–2040, 2009.

[14] M. Owais, M. Arsalan, J. Choi, and K. R. Park, "Effective diagnosis and treatment through content-based medical image retrieval (cbmir) by using artificial intelligence," *Journal of clinical medicine*, vol. 8, no. 4, p. 462, 2019.

[15] D. Zhu, T. Zhang, X. Jiang, X. Hu, H. Chen, N. Yang, J. Lv, J. Han, L. Guo, and T. Liu, "Fusing dti and fmri data: a survey of methods and applications," *NeuroImage*, vol. 102, pp. 184–191, 2014.

[16] Y. Yang, H. Chen, Z. Chang, Y. Xiang, C. Ye, and T. Ma, "Incomplete learning of multi-modal connectome for brain disorder diagnosis via modal-mixup and deep supervision," in *Medical Imaging with Deep Learning*, 2023.

[17] B. Ahmad, L. Khamidullina, A. A. Korobkov, A. Manina, J. Haueisen, and M. Haardt, "Joint model order estimation for multiple tensors with a coupled mode and applications to the joint decomposition of eeg, meg magnetometer, and gradiometer tensors," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 1186–1190.

[18] J. Dauwels, K. Srinivasan, R. M. Ramasubba, and A. Cichocki, "Multi-channel eeg compression based on matrix and tensor decompositions," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 629–632.

[19] T. Chiheb, O. Bilaniuk, D. Serdyuk *et al.*, "Deep complex networks," in *International Conference on Learning Representations. https://openreview.net/forum*, 2017.

[20] E. Grassucci, A. Zhang, and D. Comminiello, "Phnns: Lightweight neural networks via parameterized hypercomplex convolutions," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[21] A. Zalesky, A. Fornito, and E. T. Bullmore, "Network-based statistic: identifying differences in brain networks," *Neuroimage*, vol. 53, no. 4, pp. 1197–1207, 2010.

[22] 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.

[23] M. Plitt, K. A. Barnes, and A. Martin, "Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards," *NeuroImage: Clinical*, vol. 7, pp. 359–366, 2015.

[24] 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.

[25] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the abide dataset," *NeuroImage: Clinical*, vol. 17, pp. 16–23, 2018.

[26] 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.

[27] 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.

[28] 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.

[29] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 625–635.

[30] L. Peng, N. Wang, N. Dvornek, X. Zhu, and X. Li, "Fedni: Federated graph learning with network inpainting for population-based disease prediction," *IEEE Transactions on Medical Imaging*, 2022.

[31] 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.

[32] 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.

[33] L. Zhang, L. Wang, D. Zhu, A. D. N. Initiative *et al.*, "Predicting brain structural network using functional connectivity," *Medical Image Analysis*, vol. 79, p. 102463, 2022.

[34] Y. Zhang and H. Huang, "New graph-blind convolutional network for brain connectome data analysis," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 669–681.

[35] H. Hu, D. Ji, W. Gan, S. Bai, W. Wu, and J. Yan, "Class-wise dynamic graph convolution for semantic segmentation," in *European Conference on Computer Vision*. Springer, 2020, pp. 1–17.

[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. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *Acm Transactions On Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.

[38] 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.

[39] 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.

[40] R. Liu, Z.-A. Huang, Y. Hu, Z. Zhu, K.-C. Wong, and K. C. Tan, "Attention-like multimodality fusion with data augmentation for diagnosis of mental disorders using mri," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[41] Y. Huang and A. C. Chung, "Disease prediction with edge-variational graph convolutional networks," *Medical Image Analysis*, vol. 77, p. 102375, 2022.

[42] J. Ji and Y. Zhang, "Functional brain network classification based on deep graph hashing learning," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2891–2902, 2022.

[43] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 3558–3565.

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

[45] R. Saqur and K. Narasimhan, "Multimodal graph networks for compositional generalization in visual question answering," *Advances in Neural Information Processing Systems*, vol. 33, pp. 3070–3081, 2020.

[46] X. Song, F. Zhou, A. F. Frangi, J. Cao, X. Xiao, Y. Lei, T. Wang, and B. Lei, "Multi-center and multi-channel pooling gcn for early ad diagnosis based on dual-modality fused brain network," *IEEE Transactions on Medical Imaging*, 2022.

[47] 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*, 2022.

[48] T. Parcollet, M. Ravanelli, M. Morchid, G. Linarès, C. Trabelsi, R. De Mori, and Y. Bengio, "Quaternion recurrent neural networks," *arXiv preprint arXiv:1806.04418*, 2018.

[49] A. Cariow and G. Cariowa, "Fast algorithms for deep octonion networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[50] T. Chen, H. Yin, X. Zhang, Z. Huang, Y. Wang, and M. Wang, "Quaternion factorization machines: a lightweight solution to intricate feature interaction modeling," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[51] E. Grassucci, D. Comminiello, and A. Uncini, "A quaternion-valued variational autoencoder," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 3310–3314.

[52] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," *IEEE Transactions on Neural Networks and learning systems*, vol. 23, no. 4, pp. 541–551, 2012.

[53] J. Lee, S.-W. Chung, S. Kim, H.-G. Kang, and K. Sohn, "Looking into your speech: Learning cross-modal affinity for audio-visual speech separation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1336–1345.

[54] E. C. Yeats, Y. Chen, and H. Li, "Improving gradient regularization using complex-valued neural networks," in *International Conference on Machine Learning*. PMLR, 2021, pp. 11 953–11 963.

[55] Y. Quan, Y. Chen, Y. Shao, H. Teng, Y. Xu, and H. Ji, "Image denoising using complex-valued deep cnn," *Pattern Recognition*, vol. 111, p. 107639, 2021.

[56] Y. Hu, Y. Liu, S. Lv, M. Xing, S. Zhang, Y. Fu, J. Wu, B. Zhang, and L. Xie, "Dccrn: Deep complex convolution recurrent network for phase-aware speech enhancement," *arXiv preprint arXiv:2008.00264*, 2020.

[57] R. Trabelsi, I. Jabri, F. Melgani, F. Smach, N. Conci, and A. Bouallegue, "Indoor object recognition in rgbd images with complex-valued neural networks for visually-impaired people," *Neurocomputing*, vol. 330, pp. 94–103, 2019.

[58] A. Muppidi and M. Radfar, "Speech emotion recognition using quaternion convolutional neural networks," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 6309–6313.

[59] Z. Wang, X. Xu, Y. Luo, G. Wang, and Y. Yang, "Hypercomplex context guided interaction modeling for scene graph generation," *Pattern Recognition*, vol. 141, p. 109634, 2023.

[60] Y. Yang, G. Cai, C. Ye, Y. Xiang, and T. Ma, "Tensor-based complex-valued graph neural network for dynamic coupling multimodal brain networks," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[61] B. Lei, N. Cheng, A. F. Frangi, E.-L. Tan, J. Cao, P. Yang, A. Elazab, J. Du, Y. Xu, and T. Wang, "Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early alzheimer's disease," *Medical image analysis*, vol. 61, p. 101652, 2020.

[62] S. Scardapane, S. Van Vaerenbergh, A. Hussain, and A. Uncini, "Complex-valued neural networks with nonparametric activation functions," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 2, pp. 140–150, 2018.

[63] M. Wolter and A. Yao, "Complex gated recurrent neural networks," *Advances in neural information processing systems*, vol. 31, 2018.

[64] 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.

[65] C. Craddock, S. Sikka, B. Cheung, R. Khanuja, S. S. Ghosh, C. Yan, Q. Li, D. Lurie, J. Vogelstein, R. Burns *et al.*, "Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac)," *Front Neuroinform*, vol. 42, pp. 10–3389, 2013.

[66] 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.

[67] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman *et al.*, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[68] J.-D. Tournier, F. Calamante, and A. Connelly, "Mrtrix: diffusion tractography in crossing fiber regions," *International journal of imaging systems and technology*, vol. 22, no. 1, pp. 53–66, 2012.

[69] A. W. Anderson, "Measurement of fiber orientation distributions using high angular resolution diffusion imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 54, no. 5, pp. 1194–1206, 2005.

[70] R. E. Smith, J.-D. Tournier, F. Calamante, and A. Connelly, "Sift: Spherical-deconvolution informed filtering of tractograms," *Neuroimage*, vol. 67, pp. 298–312, 2013.

[71] W. Klimesch, "Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain research reviews*, vol. 29, no. 2-3, pp. 169–195, 1999.

[72] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: a user-friendly application for meg/eeg analysis," *Computational intelligence and neuroscience*, vol. 2011, pp. 1–13, 2011.

[73] B. D. Gonsalves, I. Kahn, T. Curran, K. A. Norman, and A. D. Wagner, "Memory strength and repetition suppression: multimodal imaging of medial temporal cortical contributions to recognition," *Neuron*, vol. 47, no. 5, pp. 751–761, 2005.

[74] M. A. Binnewijzend, M. M. Schoonheim, E. Sanz-Arigita, A. M. Wink, W. M. van der Flier, N. Tolboom, S. M. Adriaanse, J. S. Damoiseaux, P. Scheltens, B. N. van Berckel *et al.*, "Resting-state fmri changes in alzheimer's disease and mild cognitive impairment," *Neurobiology of aging*, vol. 33, no. 9, pp. 2018–2028, 2012.

[75] S. J. Gotts, S. C. Milleville, P. S. Bellgowan, and A. Martin, "Broad and narrow conceptual tuning in the human frontal lobes," *Cerebral Cortex*, vol. 21, no. 2, pp. 477–491, 2011.

[76] M. Grossman, P. Koenig, G. Glosser, C. DeVita, P. Moore, J. Rhee, J. Detre, D. Alsop, and J. Gee, "Neural basis for semantic memory difficulty in alzheimer's disease: an fmri study," *Brain*, vol. 126, no. 2, pp. 292–311, 2003.

[77] G. S. Dichter, "Functional magnetic resonance imaging of autism spectrum disorders," *Dialogues in clinical neuroscience*, 2022.

[78] S. Durston, P. de Zeeuw, and W. G. Staal, "Imaging genetics in adhd: a focus on cognitive control," *Neuroscience & Biobehavioral Reviews*, vol. 33, no. 5, pp. 674–689, 2009.

[79] Y. Takarae, N. J. Minshew, B. Luna, and J. A. Sweeney, "Atypical involvement of frontostriatal systems during sensorimotor control in autism," *Psychiatry Research: Neuroimaging*, vol. 156, no. 2, pp. 117–127, 2007.

[80] L. Q. Uddin, A. C. Kelly, B. B. Biswal, D. S. Margulies, Z. Shehzad, D. Shaw, M. Ghaffari, J. Rotrosen, L. A. Adler, F. X. Castellanos *et al.*, "Network homogeneity reveals decreased integrity of default-mode network in adhd," *Journal of neuroscience methods*, vol. 169, no. 1, pp. 249–254, 2008.

[81] S. Wang, Y. Yang, W. Xing, J. Chen, C. Liu, and X. Luo, "Altered neural circuits related to sustained attention and executive control in children with adhd: an event-related fmri study," *Clinical Neurophysiology*, vol. 124, no. 11, pp. 2181–2190, 2013.

[82] Z. Mao, Y. Su, G. Xu, X. Wang, Y. Huang, W. Yue, L. Sun, and N. Xiong, "Spatio-temporal deep learning method for adhd fmri classification," *Information Sciences*, vol. 499, pp. 1–11, 2019.